Author:
Carpenter, Sophie J

Title:
Induction of Specific Gene Expression in Wheat Using a dCas9-Targeted VP64 Enhancer

Induction of Specific Gene Expression in Wheat Using a dCas9-Targeted VP64 Enhancer

Sophie Jayne Carpenter

A dissertation submitted to the University of Bristol in accordance with the requirements for the award of the degree of Biology MSc by Research in the Faculty of Life Sciences.

School of Biological Sciences

July 2020

Word Count: 20,384

Abstract

Dead Clustered Regularly Interspaced Short Palindromic Repeats Associated Protein 9 (dCas9) is a catalytically dead Cas9 nuclease. It can be used in conjunction with a SunTag epitope array to recruit multiple copies of the transcriptional activator Viral Protein 64 (VP64) to any target gene, therefore increasing its expression. The system was initially developed in human cells and was shown to induce an up to 50-fold increase in gene expression. The aim of this study was to translate the technology into bread wheat, which is a vital global crop. SunTag constructs optimised for use in wheat were transiently and stably transformed into wheat protoplasts and plants respectively. The transcription, translation and interaction of these constructs were verified, although the results suggested they were weakly expressed or were deleterious in wheat. RNA sequencing (RNA-seq) analysis showed that SunTag could be used in protoplasts to increase expression of an exogenous target by 1.77-fold, however, no upregulation was detected when an endogenous gene was targeted. This study shows that the SunTag system can be implemented in wheat, however further development is necessary for it to become a useful tool. There are additional experiments which should be conducted to more comprehensively assess the power and specificity of SunTag in wheat.

Dedication and Acknowledgements

I could not have completed this project were it not for the support of some very kind people. Everyone in Lab 333 provided valuable advice at some point. Amanda Burridge proof-read numerous chapters, helped me to present my work much more effectively and answered all my tedious questions in the lab. Thank you to Jane Coghill and Christy Waterfall of the Bristol Genomics Facility for carrying out the RNA-seq analysis and to Alex Paterson and Paul Wilkinson for helping me to analyse the incredible amount of data involved in the study. A special mention goes to Christy for taking the time to proof-read a lot of this thesis. Lucy Hyde provided endless advice in the lab, dramatically improving my molecular biology skills. Thank you to Mark Winfield for teaching me how to produce wheat protoplasts and use the confocal microscope; these are skills which were a joy to learn and will be useful throughout my career. It has been a pleasure to work with Alex Coulton, Ryan Lok, Sacha Przewieslik-Allen and Calum Graham during the year. I would like to acknowledge the help of Gary Barker, who has provided advice, encouragement and a second opinion throughout the project. I would like to extend my deepest gratitude to Keith Edwards, who has dedicated a great deal of time and expertise to guiding me through this MSc. His support and encouragement have been unwavering, and have provided me with the best possible start to my research career. Finally, endless thanks go to my friends and family who have supported me through experimental breakthroughs and roadblocks alike.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

1.1 CRISPR/Cas9

Clustered regularly interspaced short palindromic repeats (CRISPRs) were first observed in *Escherichia coli* by Ishino *et al*. in 1987. Around 40% of bacteria and 90% of archaea contain CRISPRs as part of their adaptive immune systems (Makarova *et al*., 2011). Since their discovery, CRISPRs have been harnessed and modified for gene editing, resulting in a relatively quick and low-cost system known as CRISPR/Cas9 (CRISPR-associated protein 9) (Beneke *et al*., 2017; Kotwica-Rolinska *et al*., 2019).

There are two components to the CRISPR/Cas9 gene-editing system: the Cas9 nuclease and single guide RNA (sgRNA). The choice of Cas9 nuclease is important, as each demonstrates variable levels of on- and off-target activity. On-target activity, also known as editing efficiency, is defined as the percentage of target cells transformed with Cas9 in which mutations are observed in the target DNA (Raitskin *et al*., 2019). The specificity of Cas9 nucleases, which determines off-target activity, is also of vital importance as low specificity can cause a high rate of unintended mutations (Raitskin *et al*., 2019). *Streptococcus pyogenes* Cas9 (SpCas9) is the most commonly used nuclease variant, although it has been suggested that this isn't the optimum variant for gene editing purposes (Gilbert *et al*., 2013b); for instance, SaCas9 from *Staphylococcus aureus* has been shown to have significantly greater efficiency compared to SpCas9 when targeting a synthetic tobacco (*Nicotiana benthamiana*) PDS1 gene (Raitskin *et al*., 2019). Additionally, an engineered 'enhanced specificity' variant of SpCas9 has been released which has been shown to have reduced off-target effects in human cells (Slaymaker *et al*., 2016).

CRISPR RNA (crRNA) and *trans*-activating CRISPR RNA (tracrRNA) molecules were originally used to guide the Cas9 protein to a complementary target genomic DNA (gDNA) sequence. crRNA and tracrRNA molecules are now frequently combined as a chimeric sgRNA molecule (Jinek *et al*., 2012), further simplifying the system. sgRNAs contain a 20bp sequence that is complementary to the target gDNA sequence directly upstream of a protospacer adjacent motif (PAM). Every Cas9 variant has a different PAM sequence, for instance, the PAM for spCas9 is NGG (where N is any nucleotide) (Graham & Root, 2015). The sgRNA forms a complex with Cas9, guiding the nuclease to the 20bp gDNA target sequence where it introduces a double-strand break (DSB).

There are two possible pathways once the DSB has been created by the Cas9 nuclease, as detailed in Figure 1.1. Non-homologous end joining (NHEJ) is an error-prone method of repairing DSBs mediated by Ku heterodimers (Hsu *et al.*, 2014). Indels frequently occur during this pathway, causing frameshift

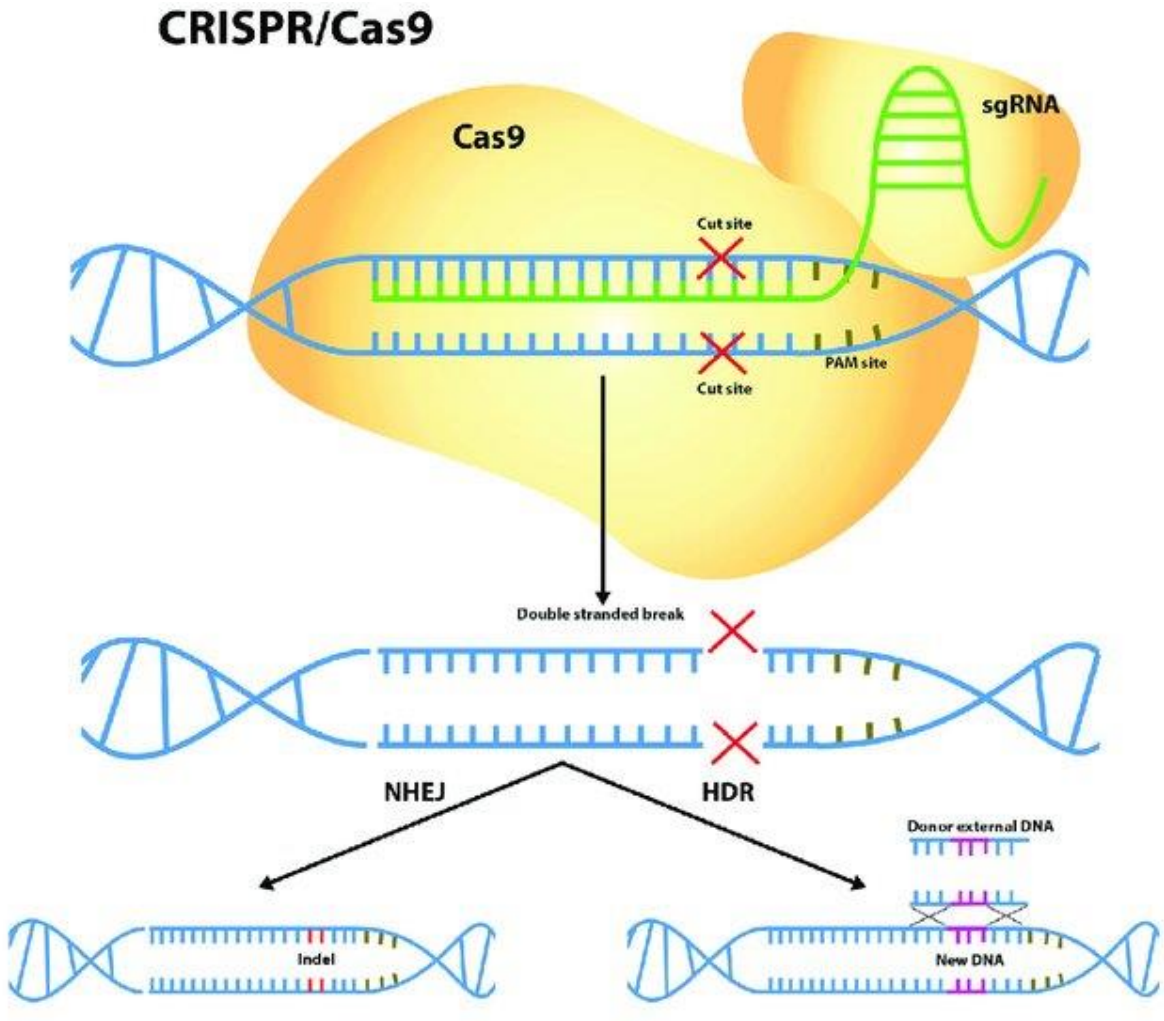

Figure 1.1: The basic CRISPR/Cas9 gene-editing mechanism. There are two gene editing pathways once a double-strand break has been created: non-homologous end-joining (NHEJ) and homology-directed repair (HDR). Reproduced from Cribbs & Perera (2017).

mutations (Hsu *et al.*, 2014). This phenomenon has been harnessed to knock-out (silence) genes of interest. Alternatively, homology-directed repair (HDR) can be used to introduce exogenous genes. Donor DNA is designed with 'homology arms' complementary to target gDNA (Zhang *et al*., 2017). When it is inserted into the target cell it is used as a template when the DSB is repaired, so inserting the donor DNA.

This novel technique has revolutionised genomics since it was first applied as a gene-editing tool in mammalian cells in 2012 (Jinek *et al*., 2012). Despite some similarities to zinc finger nucleases and TALENs (Transcription Activator-Like Effector Nucleases), which have been used for gene editing since 1996 and 2011 respectively (DeFrancesco, 2011; Kim *et al*., 1996), CRISPR/Cas9 has several advantages over either technique. The most significant of these are its adaptability and affordability (Larson *et al*., 2013). Whereas previous technologies have required complex redesign steps and

lengthy subsequent validation for novel targets (Carroll *et al*., 2006; Cermak *et al*., 2011), CRISPR/Cas9 requires only a change in the sgRNA inserted; sgRNAs are comparatively cheap and easy to design. This flexibility creates the possibility of multiplexing sgRNAs (using multiple sgRNAs in the same transformation) to either improve cleaving efficiency at a single gene or to target multiple genes concurrently.

CRISPR is increasingly the tool of choice to interrogate gene function through generating knock-out mutants and analysing the mutant phenotype (Wang *et al*., 2014b; Wang *et al*., 2018). This methodology can also be utilised in genome-wide loss-of-function genetic screens, as shown in a 2014 study by Wang *et al.* (2014a) in human cell lines. CRISPR/Cas9 avoids many disadvantages associated with previous screening techniques such as small interfering RNAs (siRNAs) (Wang *et al*., 2014a). The results of siRNA screens have been questioned as a result of significant off-target effects (Qiu *et al*., 2005; Taylor & Woodcock, 2015).

The first CRISPR-edited commercial crop (a browning-resistant button mushroom (*Agaricus bisporus*)) was approved by the US government in 2016 (Waltz, 2016), illustrating how far the technology has advanced and been accepted by the wider public in a relatively short period of time.

1.2 CRISPR ACTIVATION

The original CRISPR gene-editing system has been modified several times to unlock its full potential. This has included using alternative Cas9 nucleases such as Cas12a (also known as Cpf1) which creates a staggered DSB via a T-rich PAM (Zetsche *et al*., 2015). Furthermore, the system has been adapted to use dead Cas9 (dCas9), a catalytically 'dead' Cas9 variant with two point mutations: D10A in the RuvC1 nuclease domain and H841A in the HNH nuclease domain (Qi *et al*., 2013). These single nucleotide polymorphisms (SNPs) convey undetectable endonuclease activity. dCas9 was originally developed as an alternative knock-down technique that works by physically interfering with transcription via steric hindrance, a technique known as CRISPR interference (CRISPRi) (Larson *et al*., 2013; Qi *et al*., 2013). However, in the paper that first described CRISPRi, the potential of using dCas9 as a "flexible scaffold for directing diverse regulatory machinery to specific sites in the genome" (Qi *et al*., 2013) was noted.

This dCas9 system relies upon either fusing or recruiting effector domains to the dCas9 construct. Effector domains are the 'regulatory machinery' (Qi *et al*., 2013) mentioned previously, which affect gene expression through their effects on epigenetic states as well as transcription (Frietze & Farnham, 2011).

Epigenetic modifications are heritable traits, independent of the DNA sequence, that affect gene expression (Kagohara *et al*., 2018). There are numerous epigenetic mechanisms by which gene

expression can be altered. Chromatin structure affects the accessibility of genes to transcriptional machinery; heterochromatin has a dense structure and is transcriptionally inactive, while euchromatin is the more 'open' transcriptionally active form (Handy *et al.*, 2011). Histone proteins, which form an integral part of chromatin, can also be modified, chiefly via acetylation which generally increases gene expression (Handy *et al.*, 2011). Histone acetylation has been induced by dCas9-p300 (Hilton *et al*., 2015) and dCas9-HAT fusions (Cheng *et al*., 2016), so increasing gene expression. The addition of a methyl group to the fifth carbon of cytosines in CpG dinucleotide sequences is a common epigenetic modification which silences the gene involved (Handy *et al.*, 2011). Specific DNA methylation has been induced using dCas9-TET1 fusions (Xu *et al*., 2016) which decrease methylation (so increasing gene expression), as well as dCas9-DNMT3A fusions which increase methylation (so decreasing gene expression) (Vojta *et al*., 2016). Some effector domains have multiple mechanisms by which they induce epigenetic modifications and therefore modify gene expression. For instance, Gilbert *et al*. (2013b; 2014) fused the Krüppel-associated box (KRAB) domain from Kox1 to dCas9, which was shown to strongly repress target genes by inducing DNA methylation, histone deacetylation, histone methylation, the formation of heterochromatin, as well as steric hindrance by dCas9 itself (Gilbert *et al*., 2014; Ying *et al*., 2015).

Transcriptional activation domains (TADs) can be fused to dCas9 to increase gene expression. These can act by either interacting with transcriptional machinery directly or by interacting with endogenous transcription factors (Frietze & Farnham, 2011). The uses of CRISPR-targeted effector domains have only been partially explored; it is a rapidly expanding area of research (Figure 1.2). Examples of dCas9-TAD fusions that have been tested include dCas9-p65, dCas9-Rta and dCas9-Viral Protein 64 (VP64), of which dCas9-VP64 is the most used (Chavez *et al*., 2015; Li *et al*., 2017). VP64 is a tetramer of Viral Protein 16 (VP16), a transcription factor originating from the herpes simplex virus type 1 (Hirai *et al*., 2010). It interacts with multiple basal transcription factors (Kobayashi *et al.*, 1995; Lin *et al*., 1991; Xiao *et al*., 1994; Zhu *et al*., 1994), the Mediator complex (Ito *et al*., 1999; Mittler *et al*., 2003), as well as the cofactor PC4 (Ge & Roeder, 1994). It additionally recruits histone acetyltransferases, indirectly altering the epigenetic state of genes.

Gilbert *et al.* (2013b) fused dCas9 to several different effector domains including VP64, inducing targeted gene activation in human (*Homo sapiens*) and yeast (*Saccharomyces cerevisiae*) cells. This technique went on to be known as CRISPR activation, or CRISPRa. dCas9-transcriptional activator fusion proteins have quickly become the 'standard' technique used in CRISPRa, and have been used in a wide variety of organisms such as *E. coli* (Bikard *et al*., 2013), tobacco (Piatek *et al*., 2015) and human cells (Perez-Pinera *et al*., 2013). Despite their popularity, dCas9-VP64 fusions frequently increase gene expression less than two-fold (Tanenbaum *et al*., 2014), therefore limiting their utility.

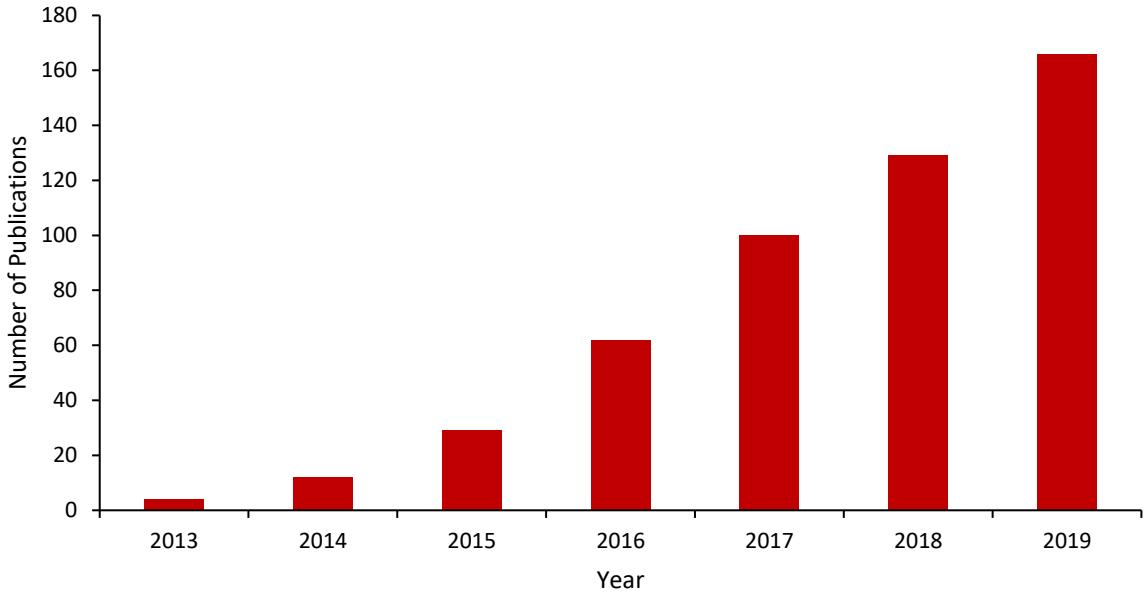

Figure 1.2: The number of publications per year resulting from a search for the topic "dCas9" on Web of Science (Clarivate Analytics, 2020b).

Second-generation systems have been developed to induce greater gene activation than simple protein fusions.

In early studies an increase in the level of activation was achieved by tiling multiple sgRNAs targeted to a single promoter (Perez-Pinera *et al*., 2013), however, the need for more than one sgRNA per target may cause issues related to the scalability of the CRISPRa system, thereby limiting its potential in overexpression screens (Konermann *et al*., 2015). Therefore, other solutions have been sought. Second-generation systems most frequently utilise effector domains other than VP64, or increase the number of effector domains recruited to the target sequence. Examples of optimised effector domains include VPR (a VP64-p65-Rta fusion), which has been shown to increase gene expression up to 320-times more potently than VP64 alone in human cells (Chavez *et al*., 2015), and TV (six copies of the TALE TAD motif from *Xanthomonas oryzae* fused to VP128), which showed up to 100-times greater gene upregulation in rice (*Oryza* spp.) compared to VP64 (Li *et al*., 2017). One limitation is that the most potent effector domain varies according to the target species, and possibly according to the target gene. For instance, while dCas9-VPR has been shown to upregulate human genes up to 20,000-fold (Chavez *et al*., 2015), a comparatively modest 46-fold increase has been observed when using dCas9-TV (Li *et al*., 2017). Therefore, it is important to undertake a comparative study of effector molecules in each target species to ensure the most potent transcriptional upregulation is achieved.

Direct fusions of high numbers of effector domains are not practical due to the corresponding increase in protein size and the issues this causes when generating transformation plasmids (Tanenbaum *et al*., 2014). Also, using multiple copies of VP64 (such as VP192 or VP256) decreases gene activation due to reduced protein production and subsequent protein stability (Li *et al*., 2017). Therefore, more complex systems are necessary to increase the number of effector domains recruited to a target promoter; these can be broadly classified into those which amplify the signal by recruiting multiple effector domain copies using the sgRNA molecule, or an epitope array fused to dCas9.

sgRNA 2.0 is a system in which multiple effector domains are recruited to a target by MS2 aptamers bound to the sgRNA molecule. This system, developed by Konermann *et al.* (2015), involves recruiting various effector domains to the same target. VP64, p65 and Heat shock factor 1 (HSF1) proved to be the most potent combination of transcriptional upregulators showing a 105-fold median improvement in upregulation compared to dCas9-VP64 across 12 genes (Konermann *et al*., 2015). This specific combination was named 'synergistic activation mediator' ('SAM') (Konermann *et al*., 2015). The study also illustrated that SAM shows highly specific gene activation with low levels of off-target activity (Konermann *et al*., 2015). Furthermore, the SAM system was used to create a 'genome-scale gene activation screen' in A375 melanoma cells to identify genes that confer BRAF inhibitor resistance (Konermann *et al*., 2015). The same group has since carried out an additional CRISPRa screen with sgRNAs targeted to long non-coding RNAs (as opposed to protein-coding genes) to identify additional BRAF inhibitor candidate genes (Joung *et al*., 2017).

As stated previously, an epitope array can be used as an alternative to the sgRNA molecule to recruit multiple effector domains and therefore amplify the gene activation signal. SunTag is one such system (Tanenbaum *et al*., 2014).

1.3 THE SUNTAG SYSTEM

The study described in this thesis is based upon the groundwork laid by Tanenbaum *et al.* (2014) in their paper titled "A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging". The paper, cited 455 times as of 18th February 2020 (Clarivate Analytics, 2020b), reported the development of a system called 'SunTag' (Tanenbaum *et al*., 2014). It involves the use of an antibody epitope array bound to dCas9 which recruits numerous copies of a single-chain antibody (ScFv). The ScFv can then be attached to another protein, for instance, Green Fluorescent Protein (GFP) (Tanenbaum *et al*., 2014). The system was named for the brightness of the fluorescent tags it could produce after SUperNova stellar explosions (Tanenbaum *et al.*, 2014). This system can be used to amplify virtually any signal. Tanenbaum *et al.* (2014) used SunTag to recruit 24 GFP ScFv

molecules to several cellular targets such as the plasma membrane; it was shown that a single SunTag complex is sufficient for fluorescent labelling (Tanenbaum *et al.*, 2014).

The group behind this work subsequently realised that this tool could be used to improve the dCas9-VP64 CRISPRa system. They fused a SunTag General Control Nonderepressible 4 (GCN_4) epitope array to dCas9 and designed a VP64 ScFv which would bind to the GCN_4 epitope, a system compared against the 'classic' dCas9-VP64 fusion protein design in Figure 1.3. In this case, 10 VP64 effector domains were recruited to a target promoter in contrast to the 24 GFP molecules recruited to targets for fluorescent imaging (Tanenbaum *et al*., 2014). Although this theoretically amplifies the activation signal to a lesser extent, Tanenbaum *et al.* (2014) found that $SunTag_{24x}$ and $SunTag_{10x}$ induced similar levels of gene activation. $SunTag_{10x}$ produced more consistent results, so was preferentially used for gene activation purposes.

This system has since been successfully implemented in multiple organisms such as human, mouse (*Mus musculus*) and fly (*Drosophila melanogaster*) cells (Chavez *et al*., 2016), as well as into plants

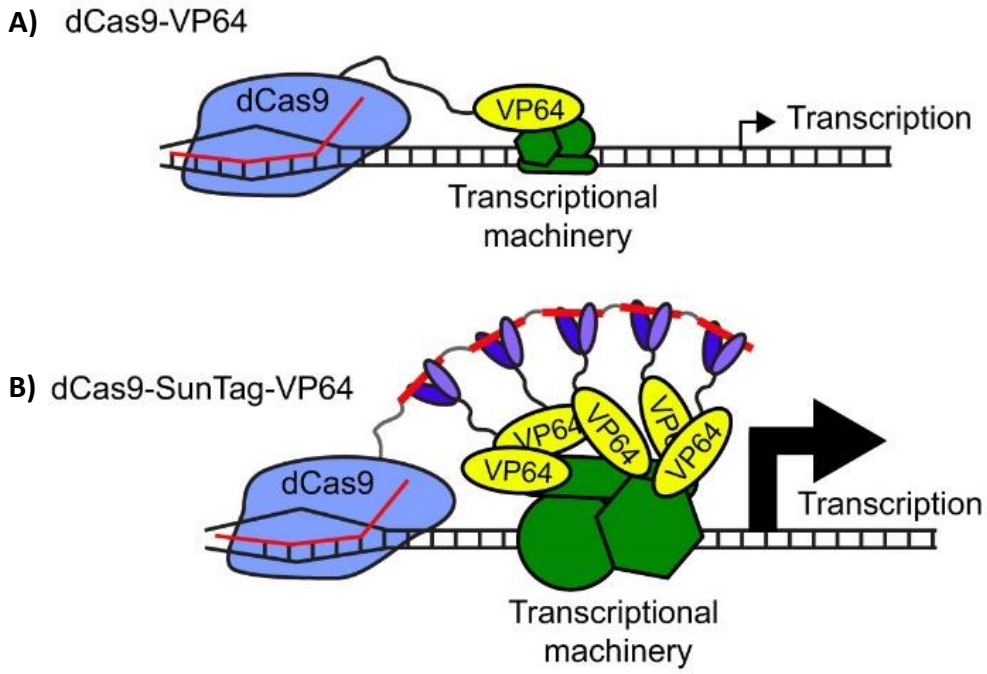

Figure 1.3: A comparison of two common CRISPRa techniques. A) The 'classic' dCas9-VP64 fusion protein, contrasted with B) the SunTag system where multiple VP64 ScFvs bind to an epitope array fused to a dCas9 protein. SunTag induces significantly higher levels of transcription than dCas9-VP64 (Tanenbaum *et al*., 2014). Reprinted from Cell, 159, Tanenbaum *et al.*, A protein-tagging system for signal amplification in gene expression and fluorescence imaging, 635-646, Copyright (2014), with permission from Elsevier (Tanenbaum *et al.* 2014).

such as *Arabidopsis thaliana* (Papikian *et al*., 2019). These are generally model organisms, so a logical next step is to translate this technology into organisms of economic importance, including crop species such as wheat.

1.4 EVOLUTIONARY HISTORY OF WHEAT

Hexaploid bread wheat (*Triticum aestivum*) is an allopolyploid resulting from multiple hybridisation events between closely related species. It contains three genomes: A, B, and D. A suggested model of the hybridisation events that occurred during the evolution of *T. aestivum* is shown in Figure 1.4.

It is generally agreed that wheat cultivation began around 10,000 years ago during the Neolithic revolution (Shewry, 2009). There is also some consensus on the identity of the diploid progenitors of the three *T. aestivum* genomes. Based on cytogenetic and repetitive sequence analysis the A genome progenitor is thought to have been *Triticum urartu*, a close relative of *Triticum monococcum* (einkorn wheat) (Dvořák *et al*., 1993; Glémin *et al*., 2019; Huang *et al*., 2002). The origin of the B genome is less certain, as an exact progenitor species is yet to be identified. An undiscovered species closely related to *Aegilops speltoides* is the most likely candidate given the similarities between the *T. aestivum* B

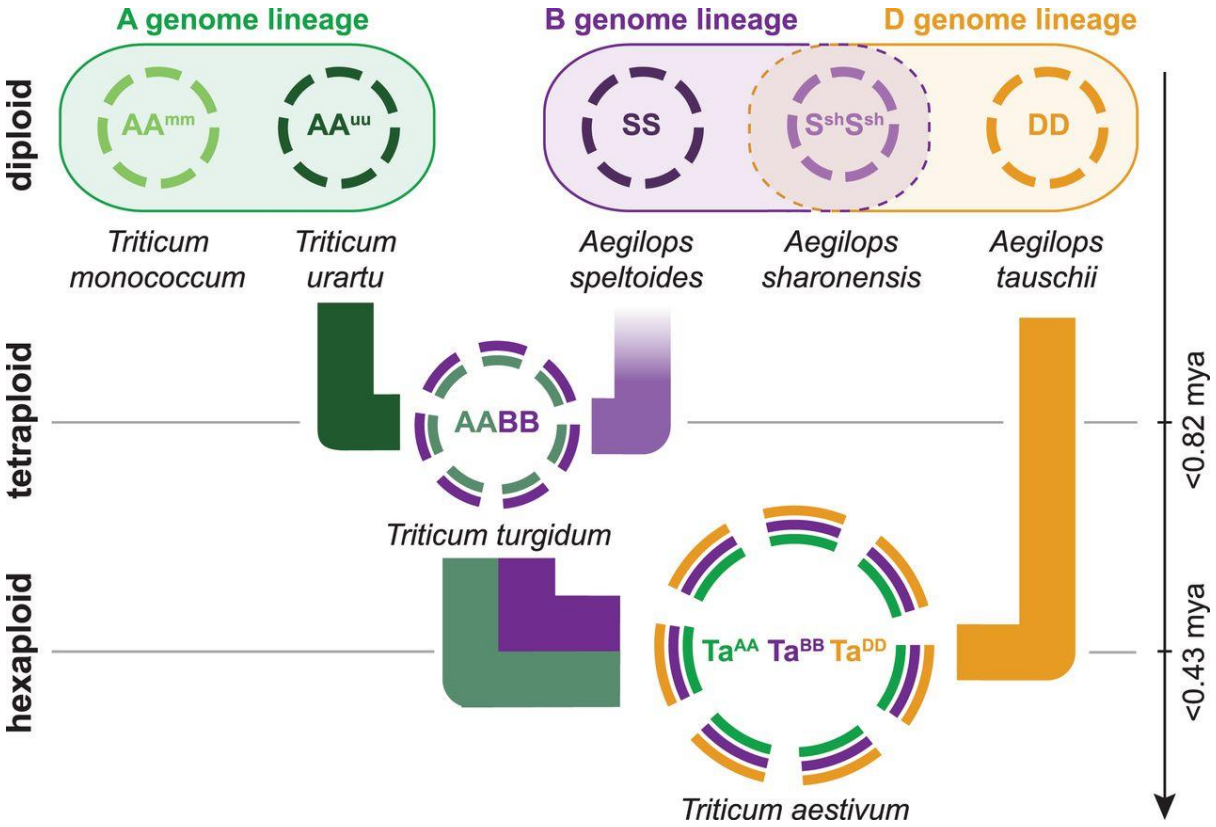

Figure 1.4: Model of the evolutionary history of bread wheat (*Triticum aestivum*; AABBDD. The chromosomal composition of each species is displayed within the circles. From IWGSC (2014). Reprinted with permission from AAAS.

genome and *A. speltoides* S genome (Glémin *et al*., 2019). However, some groups have suggested that there are multiple parental *Aegilops* progenitors of the B genome and so it is of polyphyletic origin (El Baidouri *et al*., 2017). A phylogenetic analysis of all diploid Triticeae species and a study based upon RNA sequencing (RNA-seq) data both support this multi-parental hypothesis (Petersen *et al*., 2006; Miki *et al*., 2019). Although it has been established that *Aegilops tauschii* is the D genome progenitor (Glémin *et al*., 2019; Petersen *et al*., 2006), the origin of the D genome lineage more generally is the source of significant controversy (El Baidouri *et al*., 2017). Marcussen *et al*. (2014) proposed a hypothesis where the D genome lineage arose via a single homoploid hybrid speciation event between the A and B genome lineages ~5.5 million years ago (Marcussen *et al*., 2014; Sandve *et al*., 2015). Li *et al*. (2015a; 2015b) subsequently proposed a more complex scenario involving multiple hybridisation events during the origins of the D genome. Studies have since been published in support of both hypotheses; one using amplicon sequencing of nuclear loci supported the Li *et al*. complex hybridisation hypothesis (Glémin *et al*., 2019), while another using transcriptome data supported the Marcussen *et al*. single hybridisation event hypothesis (Huynh *et al*., 2019). The origin of wheat is still a contentious issue.

The evolutionary history of wheat is relevant to this project due to the effect that polyploidy has on epigenetic states and gene expression. A tripling of gene copy number should theoretically lead to a three-fold increase in gene expression (Comai, 2005). However, this is not always the case, possibly as a result of regulatory factors which do not increase in concentration according to ploidy level, or due to epigenetic resetting (Comai, 2005). This was exemplified in a northern blot study on monoploid, diploid, triploid and tetraploid maize (*Zea mays*) plants; gene expression was found to be mostly proportional to ploidy level, however, there were some exceptions where the inverse was observed (Guo *et al*., 1996). Polyploidy can also introduce epigenetic instability (Comai, 2005). Approximately 2-2.5% of wheat genes have been estimated to have been epigenetically affected by hybridisation and polyploidisation, including regulatory changes and altered DNA methylation patterns (Comai, 2005). Polyploidy adds another level of complexity when implementing a system to induce specific gene expression in wheat, due to the effects it has on gene expression and epigenetic states. It also provides the option of targeting a single homeologue or all three.

1.5 WHEAT BREEDING PROSPECTS

Since it was first cultivated, hundreds of wheat varieties have been bred for numerous desirable traits to satisfy the growing demand for greater wheat yields. There are currently 585 wheat varieties registered on the EU database (European Commission, 2019). However, there has only been one application for authorization to grow a genetically modified (GM) wheat variety compared to 237 for

maize and eight for rice (International Service for the Acquisition of Agri-Biotech Applications, 2019); it has been described as “an orphan among genetically modified (GM) crops” (Wulff & Dhugga, 2018).

It is a vital crop, contributing 20% of calories consumed by humans globally (Food and Agriculture Organisation of the United Nations (FAO), 2017). In a move that demonstrates the importance placed on wheat as a crop, several global initiatives have recently been launched that aim to fund and coordinate wheat research. These include the G20-led Wheat Initiative (Wheat Initiative, 2016), as well as the BBSRC-funded Designing Future Wheat Institute Strategic Programme (Designing Future Wheat, 2020). As a major food source, wheat will play a significant role in ensuring future food security. It is also of political importance, with the increase in wheat prices in 2007-8 being attributed to significant civil unrest, including contributing to the Arab Spring (Shiferaw *et al*., 2013).

The human population is expected to continue growing and reach 11.8 billion by 2100 (United Nations: Department of Economic and Social Affairs: Population Division, 2017); this will require food production to increase by 2-3% per year (Hawkesford *et al*., 2013). However, since the green revolution of 1966-1985 (Pingali, 2012), wheat yields have increased by an average of just 1.6% per year (FAO, 2017); lower increases than those observed over the same period in other cereal crops (Figure 1.5).

There are several reasons for the low rate of increase in wheat yields. One is the well-documented lack of genetic diversity in elite modern varieties due to the bottleneck caused by domestication and intensive breeding efforts (Reif *et al*., 2005). This lack of diversity results in “genetic vulnerability” (Reif *et al*., 2005) to pests and disease and reduces the number of useful candidate alleles for crossbreeding. Climate change is predicted to disproportionately affect wheat compared to other major crops due to both its biology and current growing patterns (Tebaldi & Lobell, 2018). This trend has already been observed with a 15% decrease in potential yield estimated to have occurred in Russia from 1980-2008 due to climate change (Lobell *et al*., 2011). Based on Paris Agreement targets for long-term temperature increases (United Nations Framework Convention On Climate Change, 2015), wheat is predicted to see a 3.6% yield decrease by 2100, although this will likely be significantly offset by an increase in CO_2 levels (Tebaldi & Lobell, 2018).

Additionally, there are issues with the crossbreeding techniques traditionally used to introduce beneficial alleles. They are time- and resource-intensive due to wheat’s long generation time of 4-5 months (Wulff & Dhugga, 2018). Problems with sexual incompatibility and linkage drag have also been documented (Wulff & Dhugga, 2018).

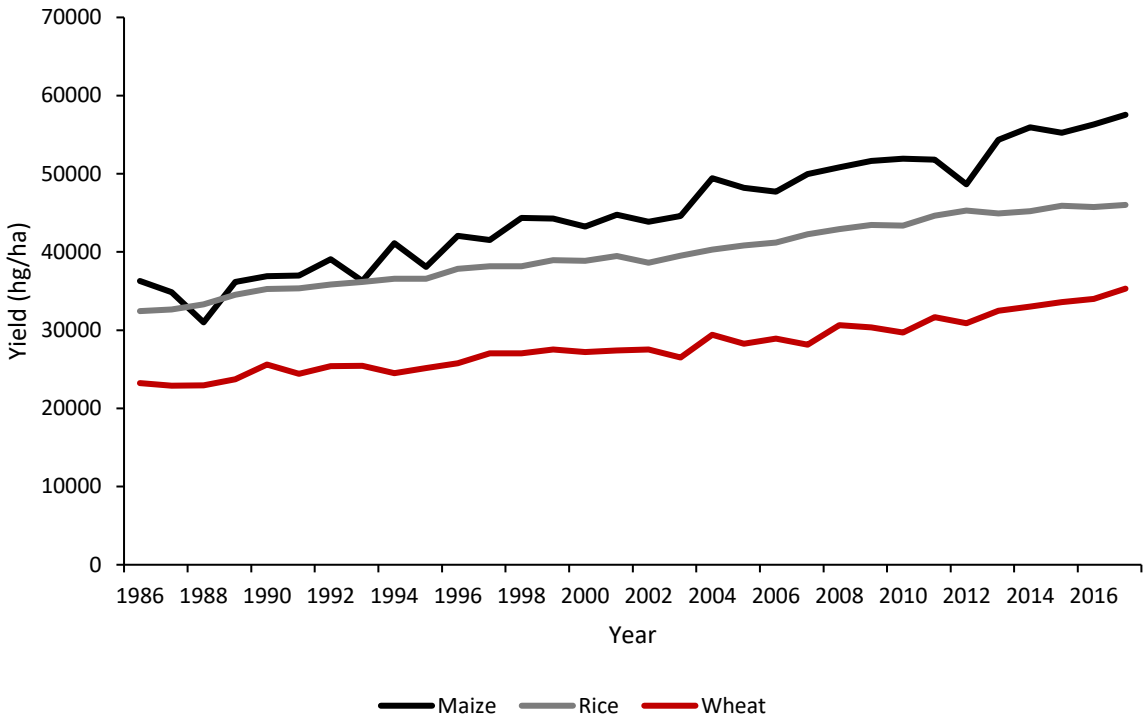

Figure 1.5: Global maize (*Zea mays*), rice (*Oryza* spp.) and wheat (*Triticum* spp.) yields from the end of the green revolution to the most recent data available (1986-2017) (FAO, 2017).

This evidence suggests that significant advances must be made to reach the required 2-3% annual increase in global yield (Hawkesford *et al*., 2013). CRISPR technology may result in significant improvements in yield through its applications to both fundamental science and crop breeding.

1.6 CRISPRa IN WHEAT

Translating tools such as CRISPRa into wheat is of vital importance due to the threats to wheat production described above. As a result of the complexity of its hexaploid genome and the fact it is a non-model organism, there has until recently been a lack of knowledge concerning this important crop species. The rate of progress has been rapid, with the landmark release of a reference genome in 2018 (IWGSC, 2018). However, there are still significant improvements to be made, as shown by the release of an improved version of the reference genome just over a year later (IWGSC, 2019).

CRISPR/Cas9 editing technology was first used in wheat in 2013 to knockout the TaMLO (Mildew resistance locus O) gene (Shan *et al*., 2013), a mutation which has been shown to confer powdery mildew resistance when all three homeologs are knocked-out (Wang *et al*., 2014b). Subsequent studies have used the technology to investigate putative grain weight genes (Zhang *et al*., 2018) as well as generate low-gluten wheat lines (Sánchez-León *et al*., 2018). Improvements have also been made to the technology itself, including delivery of the Cas9 and sgRNA constructs by

Agrobacterium tumefaciens-mediated methods as opposed to particle bombardment (Zhang *et al*., 2019). *Agrobacterium tumefaciens*-mediated transformation has been shown to reduce the occurrence of transgene silencing (Dai *et al*., 2001) as well as the number of transformed plants necessary to create a stably transformed line, although it does require a greater number of generations (Zhang *et al*., 2019). CRISPR technology has already been implemented as an effective technique in wheat, and there is huge potential to do the same with CRISPRa. Potential applications include genome-wide screens, gene function interrogation, and specific gene regulation via inducible systems. These are all valuable techniques which can be used to increase the level of basic knowledge for wheat, which will benefit breeding efforts immeasurably. This knowledge could be used to effectively target traditional crossbreeding efforts. Additionally, dependant on the future regulatory situation, CRISPRa could be used to directly create new wheat varieties with a variety of beneficial traits.

1.7 PROJECT AIMS

The broad aim of this research project is to create a system where specific gene expression can be induced in wheat using the SunTag system. This can be broken down into several parts:

1. Implement transient and stable expression of a modified SunTag system in wheat cells and plants
2. Verify that the dCas9-SunTag$_{x10}$ and VP64-ScFv SunTag constructs are present in wheat cells and plants, and are expressing the corresponding mRNA and proteins. Verify that these SunTag proteins are interacting
3. Induce specific gene expression of both inserted and endogenous genes, and quantify any changes in gene expression

The methods used to achieve these aims are described in Chapter 2. Aims 1 and 2 are broadly addressed in Chapter 3. In Chapter 4 I will describe how I assessed whether the third aim was achieved or not. Chapter 5 contains a discussion relating to the project as a whole, including its weaknesses, the impact of this study on the research area, as well as further work to be undertaken.

CHAPTER 2: MATERIALS AND METHODS

All chemicals used were supplied by Sigma-Aldrich Corporation (St. Louis, MO, USA) unless otherwise stated.

2.1 TRANSFORMATION CONSTRUCTS

For the SunTag system to function, two proteins are required: dCas9-SunTag$_{x10}$ and VP64-ScFv. The constructs used in this study were based on the plasmids published by Tanenbaum *et al*. (2014) (Addgene IDs 60904 and 60903). Various alterations were made to these constructs by Professor Keith Edwards (University of Bristol; Bristol, UK) to optimise them for use in wheat.

Figure 2.1 contains schematic representations of the constructs used throughout this study.

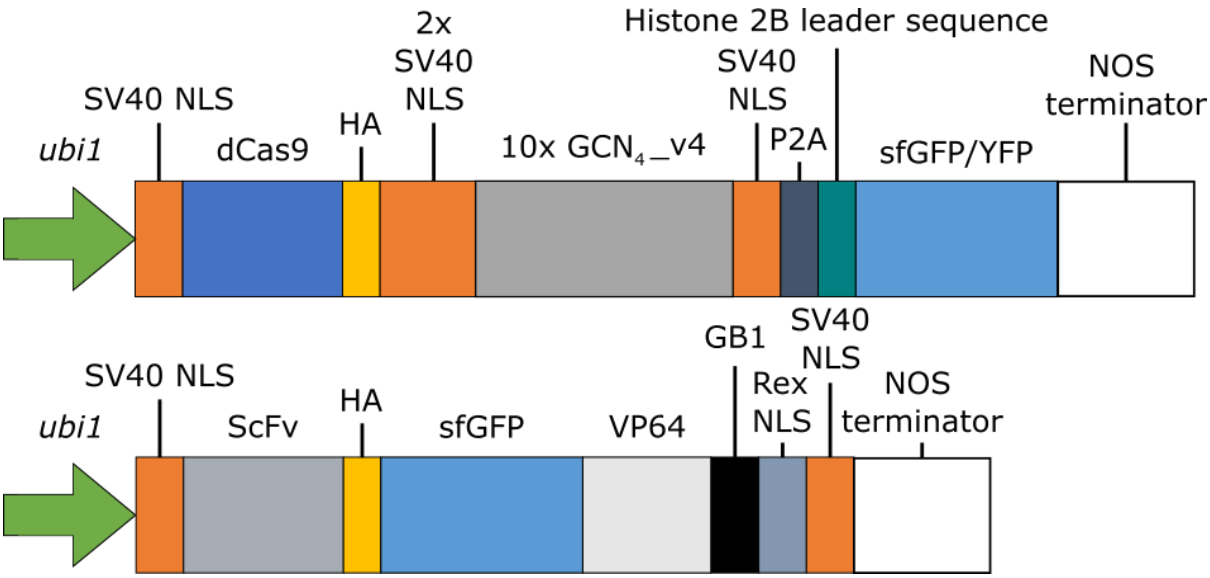

Figure 2.1: Schematic representation of the wheat-optimised A) dCas9-SunTag$_{x10}$ and B) VP64-ScFv constructs used throughout this study. *ubi1* represents the maize *ubiquitin1* promoter region. NLS stands for nuclear localisation signal; the constructs contain NLSs from simian vacuolating virus 40 (SV40) and human T-cell leukaemia virus 1 (Rex). HA represents a human influenza hemagglutinin tag. 10x GCN$_4$_v4 represents 10 copies of a 19 amino acid region of the General Control Nonderepressible protein 4 (GCN$_4$) from *S. cerevisiae* (Harmansa & Affolter, 2018). P2A is a self-cleaving motif from porcine teschovirus-1. sfGFP represents superfolder green fluorescent protein (Pédelacq *et al*., 2006). The histone 2B leader sequence is a nuclear localisation sequence from *A. thaliana*. ScFv represents a single chain variable fragment specific to the GCN$_4$ epitope. VP64 represents a tetramer of the VP16 (Viral Protein 16) activation domain from the herpes simplex virus. GB1 is the protein G B1 solubilisation domain from *Streptococcus* sp.. The NOS terminator represents the *A. tumefaciens* "Nopaline Synthase" terminator.

2.2 TRANSIENT TRANSFORMATION

2.2.1 Transformation Plasmids

Figures A.1, A.2, and A.4 contain maps of the plasmids used for transient protoplast transformation. The constructs outlined in Figure 2.1 were ligated into separate Gateway™ pENTR™4 dual selection plasmid vectors (Invitrogen; Carlsbad, CA, USA) which were subsequently used to transform competent *E. coli* cells by Professor Keith Edwards (University of Bristol; Bristol, UK). The colonies were provided as 50% v/v ethylene glycol overnight culture solutions. The cultures were streaked on kanamycin (50 mg/ml) lysogeny broth (LB) (1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v sodium chloride) agar plates and grown at 42 °C for approximately 16 h. The inoculated kanamycin plates were then stored long-term at 4 °C. A transformed *E. coli* colony was transferred to 50 ml LB containing 25 µg/ml kanamycin and incubated for approximately 7 h at 37 °C, 200 RPM. This culture was transferred to 500 ml LB also containing 25 µg/ml kanamycin and incubated at 37 °C, 200 RPM for approximately 16 h. Plasmid DNA was subsequently isolated from this culture using the QIAGEN (Hilden, Germany) Plasmid Mega Kit according to the manufacturer's instructions. The yield of plasmid DNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific; Waltham, MA, USA) using the nucleic acid 'DNA-50' setting. Plasmids were stored in elution buffer (EB) (QIAGEN) at 1 µg/µl at −20 °C.

All plasmid preparations underwent basic quality control using gel electrophoresis to detect possible nucleic acid contamination. An appropriate mass of agarose powder (Thermo Fisher Scientific) was added to TAE buffer (Thermo Fisher Scientific) to make a % w/v solution appropriate for the expected fragment sizes, based on a guide published by Thermo Fisher Scientific (Thermo Fisher Scientific, 2012a). This was heated in a microwave until fully dissolved. The solution was allowed to cool slightly, and ethidium bromide was added to a final concentration of 1.69 µM. The solution was then stored at 65 °C until needed. Once the gel was poured and allowed to set, 10 µl of appropriately diluted sample and 3 µl of loading buffer (0.62 mM bromophenol blue, 752.49 mM glycerol) was loaded per lane, with 5-13 µl of 1 kb+ DNA ladder (New England Biolabs Inc.; Ipswich, MA, USA) in flanking lanes. Gels were run at 80-120 V until fragments were fully separated (1-3 h) and visualised using a GelDoc-It TS2 Imager (UVP, LLC; Upland, CA, USA).

Sanger sequencing was used to confirm the plasmids contained the correct constructs. Polymerase chain reactions (PCRs) were carried out on the plasmid preparations to amplify the regions containing the SunTag constructs. Primer sequences are outlined in Table A.3. Per reaction, 20 pmol (2 µl) of forward primer, 20 pmol (2 µl) of reverse primer, 17 µl of sterile distilled water, 25 µl of HotStarTaq Plus master mix (2x) (QIAGEN) and 4 µg of template DNA was used. The thermocycler (Eppendorf

(Hamburg, Germany) Mastercycler® Nexus Gradient) conditions were as recommended by the master mix manufacturer.

Sequencing reactions were carried out using the BigDye™ terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions, except that the incubation step during cycle sequencing was carried out at 94 °C for 30 s as opposed to the recommended 96 °C for 1 min and, per sequencing reaction, 3 µl of purified PCR product and 3 µl of sterile distilled water were used. Primer sequences are outlined in Table A.3.

Once the sequencing reaction had been carried out, the products underwent ethanol/EDTA precipitation. Per reaction product 10 µl sterile distilled water, 5 µl 125 mM EDTA (VWR International; Radnor, PA, USA) and 60 µl of 100% v/v ethanol (VWR International) were added and mixed by vortexing. The samples were centrifuged in an Eppendorf MiniSpin Plus at 11,337 x *g* for 10 min. The supernatant was removed, 100 µl 75% v/v ethanol was added, and the samples were centrifuged in an Eppendorf MiniSpin Plus at 11,337 x *g* for 10 min. The supernatant was removed by gentle aspiration, and the samples dried in a vacuum (Thermo Scientific (Waltham, MA, USA) Savant™ SPD1010 SpeedVac™) (vacuum level: 680 Pa) for 20 min. The pellet was resuspended in 15 µl Hi-Di™ formamide (Applied Biosystems) by pipetting up and down. The samples were loaded onto a 96-well plate, and all air bubbles removed by centrifuging on the 'short' setting of an Eppendorf 5804 R centrifuge. The plate was then loaded onto a 3500 Genetic Analyzer (Applied Biosystems). The resulting sequences were quality checked and trimmed using Chromas v2.6.6 (Technelysium Pty Ltd, 2018) and aligned to the expected plasmid sequence using Sequencher v5.4.6 (Gene Codes Corporation, 2017).

2.2.2 Protoplast Isolation and PEG-Mediated Transformation

The following wheat protoplasting protocol was based on that published by Shan *et al.* (2014). Approximately 20 shoots were cut using scissors from 9-12 d old *Triticum aestivum* cv. Cadenza seedlings grown using Levington Advance Seed and Modular F2 compost (Israel Chemicals Ltd.; Tel Aviv, Israel) at 25 °C in the dark. These were washed in distilled water and cut into 0.5-1 mm latitudinal strips using a razor blade. The shoots were incubated in plasmolysis buffer (600 mM D-mannitol) at room temperature in the dark for 10 min. The shoots were then transferred into 20 ml enzyme solution (50 mM potassium chloride (Mettler-Toledo International Inc.; Columbus, OH, USA), 600 mM D-mannitol, 20 mM MES (Jena Bioscience; Jena, Germany), 1.5% w/v cellulase Onozuka RS from *Trichoderma viride* (SERVA Electrophoresis GmbH; Heidelberg, Germany), 0.75% w/v macerozyme R-10 from *Rhizopus* sp. (SERVA Electrophoresis GmbH), 10 mM calcium chloride (VWR International), 0.1% w/v bovine serum albumin (BSA)) and swirled gently to mix. The solution

was then vacuum infiltrated at room temperature in the dark in a UNIVAPO 150 vacuum concentrator (UniEquip Laborgerätebauund Vertriebs GmbH; Planegg, Germany) using a UNIJET II refrigerated aspirator (Uniequip Laborgerätebauund Vertriebs GmbH) for 30 min. Following this, the solution was incubated in the dark at 25 °C, 40 RPM for 4 h-overnight. After the incubation, 49.54 ml of room temperature W5 solution (126.16 mM calcium chloride, 155.43 mM sodium chloride (Invitrogen), 100 mM MES, 5.37 mM potassium chloride) was added to the beaker and poured through a 40 µm sterile cell strainer (Fisher Scientific; Loughborough, UK) into two 50 ml falcon tubes. The samples were centrifuged at 80 x *g* in an Eppendorf 5804 R for 3 min at 4 °C. The supernatant was discarded, and the pellets were resuspended in 3 ml of ice-cold W5 solution and pooled in a single tube. A sample was removed for cell counting using a haemocytometer (Weber Scientific; Hamilton, NJ, USA) with an H-shaped moat under a light microscope (Olympus Corporation; Tokyo, Japan) at 10x magnification. The cells in four 0.1 µl squares were counted and the mean was used to estimate the total number of cells in the sample. The remaining protoplast cells were incubated on ice for 30 min, centrifuged at 80 x *g* in an Eppendorf 5804 R for 1 min at 4 °C and the supernatant was removed. The protoplasts were gently resuspended in MMG (400 mM D-mannitol, 15 mM magnesium chloride (Invitrogen), 4 mM MES) at a concentration of 1 x 10^6 cells per ml.

Per transformation, 10 µg of plasmid, 100,000 protoplasts, and 110 µl PEG-CTS (200 mM D-mannitol, 100 mM calcium chloride, 40% w/v polyethylene glycol (PEG) 4000 (PanReac AppliChem ITW Reagents; Chicago, IL, USA)) was used. A GFP reporter plasmid (Figure A.1) was used as a positive control. The transformation reactions were incubated in the dark for 10 min at room temperature. The transformation was stopped by adding 880 µl room temperature W5 solution to each reaction and mixed by inverting the tubes. The samples were centrifuged at 81 x *g* in an Eppendorf MiniSpin Plus for 3 min at room temperature. The supernatant was removed, and the protoplasts were resuspended in 1.8 ml room temperature W5 solution. The protoplasts were incubated in the dark at 23 °C for 24-48 h.

2.2.3 Protoplast Genotyping (Fluorescent Microscopy)

A Leica Camera AG (Wetzlar, Germany) DM2000 microscope and Leica Camera AG SFL4000 illumination source were used to identify transformed cells via the nuclear-targeted GFP both transformation constructs were tagged with. Excitation was set at 470 nm and 20%. Brightfield, GFP, and blue-green-red (BGR) filter cubes were used, and Leica application suite v4.4 software (Leica Microsystems, n.d.) was used to capture images.

2.3 STABLE TRANSFORMATION

2.3.1 Transformation by Particle Bombardment

Transformation of hexaploid bread wheat cv. Cadenza callus was carried out using simultaneous particle bombardment using the constructs described in section 2.1. This was performed by Caroline Sparks (Rothamsted Research; Harpenden, UK).

Experiments were mostly conducted using plants from the second generation (T2) B3598 R2 P1 line (referred to as B3598 hereafter). This line was shown by genotyping (as described in section 2.3.2) to be segregating for both SunTag constructs. Some studies were carried out on plants from the E3 (T3) line, which again was segregating for both SunTag constructs. All individuals were grown at 18 °C in 16 h light, 8 h dark cycles. Levington Advance Seed and Modular F2 compost was used.

2.3.2 Transformed Plant Line Genotyping

Total DNA was isolated from B3598 and E3 leaf tissue harvested approximately two weeks after germination. Sections of leaf tissue approximately 2 cm in length were collected from all plants and placed inside 2 ml microcentrifuge tubes along with two 2 mm ball bearings and all samples were frozen for a minimum of 2 h at −80 °C. These samples were homogenised in a Geno/Grinder 2000 (SPEX® SamplePrep; Metuchen, NJ, USA) at 1x rate for 3 min. Per sample 600 µl extraction buffer (0.1 M tris hydrochloride (pH 7.5), 50 mM EDTA (pH 8.0), 50 mM sodium chloride, 1.25% v/v SDS (Invitrogen)) preheated to 55 °C was added and shaken vigorously for a few seconds. The samples were incubated at 55 °C for 20 min, followed by 4 °C for 5 min. To each sample, 300 µl 6 M ammonium acetate (pre-chilled to 4 °C) was added. The tubes were again shaken vigorously for a few seconds. The samples were then incubated at 4 °C for 15 min. The samples were centrifuged at 14,104 x *g* in an Eppendorf MiniSpin Plus for 15 min at room temperature to pellet the precipitated proteins and other plant tissue. The supernatant was recovered (approximately 600 µl) and transferred to new 1.5 ml microcentrifuge tubes containing 360 µl isopropanol (Acros Organics BVBA; Geel, Belgium) and mixed by inverting the tubes several times. The samples were incubated at room temperature for 5 min and then centrifuged at 14,104 x *g* in an Eppendorf MiniSpin Plus for 15 min at room temperature to pellet the DNA. The supernatant was removed, and the DNA pellet was washed in 400 µl 70% v/v ethanol and centrifuged at 14,104 x *g* in an Eppendorf MiniSpin Plus for 5 min at room temperature. The supernatant was removed. The DNA pellet was resuspended in 100 µl EB. If it was difficult to resuspend the pellet, the sample was vortexed and heated to 50 °C for 10 min in a block heater and vortexed a second time. The concentration of the samples was quantified using a Nanodrop 1000 spectrophotometer as described in section 2.2.1. DNA samples were stored at −20 °C.

To identify which plants contained construct DNA, PCR reactions were carried out as described in section 2.2.1. Gel electrophoresis was carried out on the PCR products as described in section 2.2.1.

2.4 TRANSCRIPTOMIC ANALYSIS

2.4.1 RNA Extraction

Total RNA was isolated from protoplasts, B3598 leaf tissue, and E3 leaf tissue using a TRIzol™-based method. In the case of protoplasts, the cells were pelleted by centrifugation at 43 x *g* in an Eppendorf MiniSpin Plus for 2 min at room temperature and the supernatant was removed, being careful not to disturb the loose pellet. In the case of the leaf tissue extractions, 2 cm of leaf tissue was collected from all plants and placed inside 2 ml microcentrifuge tubes along with two ball bearings and all samples were frozen for a minimum of 2 h at −80 °C. These samples were subsequently homogenised in a Geno/Grinder 2000 at a 1x rate for 3 min. To each sample, 1 ml of TRIzol™ reagent (Invitrogen) was added and vortexed for 10 s. The samples were incubated at room temperature for 2-3 min, after which 200 µl of chloroform was added to each. The microcentrifuge tubes were shaken vigorously for 15 s each and incubated at room temperature for 2-3 min. The samples were centrifuged at 1699 x *g* in an Eppendorf 5804 R for 20 min at 4 °C. For the remainder of the protocol, tubes were kept on ice. The colourless aqueous phase (approximately 1 ml) was transferred to new 15 ml falcon tubes and 500 µl isopropanol was added. The samples were mixed by inverting the tubes and centrifuged at 1699 x *g* in an Eppendorf 5804 R for 30 min at 4 °C and the supernatant was removed. To wash the pellet, 5 ml of 75% v/v ethanol was added, and the samples were centrifuged at 1699 x *g* in an Eppendorf 5804 R for 30 min at 4 °C. As much supernatant as possible was removed and the pellet was air-dried. The RNA pellet was finally resuspended in 100 µl of sterile distilled water (Invitrogen) by pipetting up and down. To eliminate any possible DNA contamination, 87.5 µl of the RNA sample was added to 10 µl RDD buffer (QIAGEN) and 2.5 µl DNAase I stock solution (Ambion, Inc; Austin, TX, USA) and incubated at room temperature for 30 min. RNA sample clean-up was then carried out using an RNeasy MinElute cleanup kit (QIAGEN) according to the manufacturer's instructions, except for step 5 where the centrifugation time was reduced from 2 min to 60 s, and step 7 where 25 µl of RNAse-free water was used to elute the RNA instead of the recommended 14 µl. The concentration of the samples was quantified using a Nanodrop 1000 spectrophotometer using the nucleic acid 'RNA-40' setting. RNA samples were stored at −80 °C.

2.4.2 RT-PCR

Reverse transcriptase-PCRs (RT-PCRs) were carried out using a OneStep RT-PCR kit (QIAGEN) according to the manufacturer's instructions. Primer sequences are outlined in Table A.3. Gel electrophoresis of

the RT-PCR products was carried out as described in section 2.2.1. Sanger sequencing of the RT-PCR products was carried out as described in section 2.2.1.

2.5 PROTEOMIC ANALYSIS

2.5.1 Protein Extraction

Total denatured protein was isolated from transformed protoplasts, B3598 leaf tissue, and E3 leaf tissue. In the case of protoplasts, the cells were pelleted by centrifugation at 43 x *g* in an Eppendorf MiniSpin Plus for 2 min at room temperature, and most of the supernatant was removed being careful not to disturb the loose pellet. In the case of the leaf tissue extractions, 2 cm of leaf tissue was collected from all plants and placed inside 2 ml microcentrifuge tubes along with two ball bearings and all samples were frozen for a minimum of 2 h at −80 °C. These samples were subsequently homogenised in a Geno/Grinder 2000 for 3 min. The homogenised plant material was suspended in 1 ml ice-cold methanol with 0.2% v/v protease inhibitor cocktail, mixed by vortexing, and incubated at −20 °C for 5 min. The samples were centrifuged at 16,000 x *g* in an Eppendorf 5804 R for 5 min at 4 °C and the supernatant was removed. This methanol wash was carried out a total of four times. The pellet was then resuspended in 1.5 ml acetone pre-chilled to −20 °C. The samples were incubated at −20 °C for 5 min, then centrifuged at 16,000 x *g* in an Eppendorf 5804 R for 5 min at 4 °C. The supernatant was removed, and the pellet was air-dried at room temperature for 10 min. The pellet was suspended in 200 µl of Reagent Type 4 Working Solution (Protein Extraction Reagent Type 4, 1% v/v protease inhibitor cocktail) by vortexing for 15 min at room temperature. The samples were then centrifuged at 16,000 x *g* in an Eppendorf 5804 R for 30 min at room temperature. The supernatant was collected and transferred to a fresh 1.5 ml microcentrifuge tube. The concentration of the samples was quantified using a Bradford Assay (Bradford, 1976). All protein samples were diluted by a factor of 10 to ensure the buffer was compatible with the Bradford reagent. The Bradford assay was carried out according to the standard 3.1 ml assay protocol recommended by the Bradford reagent manufacturer. BSA was used to prepare protein standards. All assays were set up in triplicate. Protein concentration was quantified using the Nanodrop 1000 spectrophotometer using the 'Protein Bradford' setting. Protein samples were stored at −20 °C.

2.5.2 Preparation of Cell Lysate

This protocol was based on protocols published by Abcam (2010) and Shan *et al*. (2014). Cell lysate was isolated from B3598 leaf tissue or protoplasts. In the case of protoplasts, the cells were pelleted by centrifugation at 12,000 x *g* in an Eppendorf MiniSpin Plus for 2 min at room temperature, and most of the supernatant was removed being careful not to disturb the loose pellet. The samples were incubated with 10 µl of ice-cold RIPA buffer (25 mM tris hydrochloride (pH 7.6), 50 mM

sodium chloride, 1% v/v NP-40, 1% w/v sodium deoxycholate, 0.1% w/v SDS, 1% v/v protease inhibitor cocktail) on ice for 1 min. The samples were then incubated at 4 °C on a rocker for 30 min. Following this, the samples were centrifuged at 16,000 x *g* in an Eppendorf 5804 R for 20 min at 4 °C, and the supernatant was transferred to chilled 1.5 ml microcentrifuge tubes.

In the case of the leaf tissue extractions, 2 cm of leaf tissue was collected from all plants and placed inside 2 ml microcentrifuge tubes along with two ball bearings and all samples were flash-frozen in liquid nitrogen and stored at −80 °C. These samples were subsequently homogenised in a Geno/Grinder 2000 at a 1x rate for 3 min. Per sample, 600 µl ice-cold RIPA buffer was added, mixed by vortexing, and incubated on a rocker for 2 h at 4 °C. The samples were centrifuged at 17,172 x *g* in an Eppendorf 5804 R for 20 min at 4 °C, and the supernatant was transferred to chilled 1.5 ml microcentrifuge tubes.

The concentration of the samples was quantified using a Bradford Assay as described in section 2.5.1. Protein samples were stored at −20 °C.

2.5.3 Protein Gel Electrophoresis

To reduce the samples, 10 µl of protein sample was combined with 5 µl of NuPAGE™ LDS sample buffer (4x) (Invitrogen) and 1 µl of NuPAGE™ sample reducing agent (10x) (Invitrogen), then incubated at 70 °C for 10 min. The samples were allowed to cool to room temperature for 1-2 min. A NuPAGE™ 3-8% tris-acetate protein gel (Invitrogen) was set up within an XCell SureLock™ Mini-Cell (Invitrogen). The running buffer used consisted of 5% v/v NuPAGE™ MOPS SDS running buffer (20x) (Invitrogen) and 0.1% v/v NuPAGE™ antioxidant (Invitrogen). The samples were loaded onto the gel along with 15 µl of SeeBlue™ Plus 2 pre-stained protein standard (Invitrogen) in the first lane. Cas9 nuclease (New England BioLabs Inc.) was included as a positive control where possible. Each gel was electrophorized at 150 V for 45-60 min.

2.5.4 Coomassie Blue Staining

The electrophorized gel was removed from the plastic cassette and rinsed in distilled water. Enough SimplyBlue™ SafeStain Coomassie® G-250 stain (Invitrogen) was poured over the gel to cover it and it was incubated on a rocker for 1 h at room temperature. The gel was then rinsed with distilled water on a rocker for 30 min at room temperature 2-3 times. Finally, it was rinsed with distilled water overnight at room temperature and imaged using a Huawei Technologies Co., Ltd. (Shenzhen, China) P20 mobile phone camera.

2.5.5 Western Blotting

Western blotting was carried out on electrophorized gels that had not undergone Coomassie blue staining. The transfer buffer consisted of 5% v/v NuPAGE transfer buffer (20x) (Invitrogen), 0.1% v/v NuPAGE™ antioxidant and 10% v/v methanol. Blotting pads from the XCell II™ blot module (Invitrogen) were saturated with transfer buffer. An Invitrolon™ 0.45 µm polyvinylidene fluoride (PVDF) membrane (Invitrogen) was pre-wet in methanol, rinsed in distilled water and submerged in transfer buffer for 2-3 min. The filter paper from the Invitrolon™ sandwich (Invitrogen) was submerged briefly in transfer buffer immediately before use. The electrophorized gel was removed from the plastic cassette, and the wells and lip were removed using a razor blade. The blot module was assembled according to the manufacturer's instructions, placed in a SureLock™ Mini-Cell and filled with transfer buffer so the blotting pads were just covered. The outer chamber was filled with 650 ml of distilled water. The blot was run at 30 V (220-180 mA) for 60-90 min. The blot was probed using the WesternBreeze™ chemiluminescent kit (anti-rabbit) (Invitrogen) according to the manufacturer's instructions. Cas9 antibody 3 (antibody details are contained in Table A.1) was used as the primary antibody at the manufacturer recommended concentration of 1 µg/ml. The blot was imaged with a Vilber Lourmat Sté (Collégien, France) Fusion Pulse using the ECLAppStd setting and the recommended exposure length.

2.5.6 Co-Immunoprecipitation

Cell lysate was obtained from plant B3598.6. Eight 2 cm pieces of leaf tissue were snap-frozen in 2 ml microcentrifuge tubes along with two ball bearings in liquid nitrogen. All samples were frozen for a minimum of 2 h at −80 °C then homogenised in a Geno/Grinder 2000 at 1x rate for 3 min. To each sample, 500 µl IP lysis/wash buffer from the Pierce™ co-immunoprecipitation kit (Thermo Scientific) was added and incubated on a rocker at 4 °C for 1 h. The samples were then centrifuged at 13,000 x *g* in an Eppendorf 5804 R for 10 min at 4 °C, and the supernatants were pooled in a single 5 ml microcentrifuge tube. Co-immunoprecipitation reactions were carried out using the Pierce™ co-immunoprecipitation kit according to the manufacturer's instructions, including pre-clearing of the cell lysate and use of the conditioning buffer. Five sets of agarose beads were set up in spin columns with the following treatments:

1. 75 µg Cas9 antibody 1
2. 75 µg Cas9 antibody 2
3. 75 µg Cas9 antibody 3
4. Quenched antibody coupling resin (control 1)
5. Control resin (control 2)

The protein concentration of the flow-through during the coupling reaction was measured using a Nanodrop 1000 spectrophotometer using the Protein A280 '1 Abs = 1 mg/ml' setting. Co-immunoprecipitation reactions were carried out on 500 µl of pre-cleared cell lysate per spin column and the protein concentration of the eluted product was estimated using the Pierce™ BCA protein assay kit (Thermo Scientific) according to the manufacturer's instructions, using the microplate procedure with a sample to working reagent (WR) ratio of 1:20. The Nanodrop 1000 spectrophotometer was used to quantify protein concentration using the Protein BCA setting.

2.5.7 Protein Digestion

Protein digestion was carried out by Dr Kate Heesom (University of Bristol Proteomics Facility; Bristol, UK). The samples were reduced by adding 10 mM tris(2-carboxyethyl)phosphine (TCEP) and incubating at 55 °C for 1 h. The samples were then alkylated by incubating them with 18.75 mM iodoacetamide at room temperature for 30 min. Precipitation was carried out overnight by adding acetone to a final concentration of 85.71% v/v. Following this, the samples were centrifuged at 8000 x *g* for 10 min and the supernatant was removed. Proteins were resuspended in 50 mM tetraethylammonium bromide (TEAB) and digested using 5% w/w grade trypsin (Promega UK; Southampton, UK) overnight at 37 °C.

2.5.8 Liquid Chromatography-Mass Spectrometry (LC-MS)

LC-MS was carried out by Dr Kate Heesom (University of Bristol Proteomics Facility; Bristol, UK) on peptides resulting from the protein digestion described in section 2.5.7. The peptides were desalted using a Sep Pak cartridge (Waters Corporation; Milford, MA, USA) according to the manufacturer's instructions. The eluate from the Sep Pak cartridge was evaporated to dryness and resuspended in 1% v/v formic acid. Fractionation was carried out on the peptides using an Ultimate 3000 nano-LC system and LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). The peptides were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific) and 0.5% v/v acetonitrile was used to wash the peptides. Peptides dissolved in 0.1% v/v formic acid were then resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient. The concentration of solvent A (80% v/v acetonitrile in 0.1% v/v formic acid) in 0.1% v/v formic acid was altered over seven gradient segments (flow rate: 300 nl/min):

1. 1-6% v/v solvent A over 1 min
2. 6-15% v/v solvent A over 58 min
3. 15-32% v/v solvent A over 58 min
4. 32-40% v/v solvent A over 5 min
5. 40-90% v/v solvent A over 1 min

6. 90% v/v solvent A for 6 min
7. 90-1% v/v solvent A over 1 min

Peptides were ionized by nano-electrospray using a stainless-steel emitter (Thermo Scientific) (voltage: 2.1 kV; internal diameter: 30 µm; capillary temperature: 250 °C). Tandem mass spectra were acquired using an LTQ-Orbitrap Velos mass spectrometer controlled by Xcalibur 2.1 software (Thermo Fisher Scientific, 2010). The settings used for mass spectrometry are detailed in the appendix (section A.4).

The resulting raw data files were analysed using Proteome Discoverer software v1.4 (Thermo Fisher Scientific, 2012b). Peptides detected in the samples were searched against the UniProt *T. aestivum* protein database (The UniProt Consortium, 2019) (142,582 amino acid sequences) in addition to dCas9-SunTag$_{x10}$, VP64-ScFv and GFP protein sequences. The SEQUEST algorithm (Eng *et al*., 1994) was used for the search, and the parameters used are detailed in the appendix (section A.5)

2.6 RNA-SEQ

2.6.1 Sample Preparation

Protoplasts were isolated from both wild-type cv. Cadenza (WT) and B3598.6 seedlings and transformed according to the protocol outlined in section 2.2.2. Details of the sgRNAs used are in Table A.2 and Figure 2.2. Six transformation treatments were applied to protoplasts in duplicate as outlined in Table 2.1.

Table 2.1: Details of the treatments applied to batches of protoplasts for RNA-seq analysis. The protoplasts were transformed with pools of plasmids containing one of the appropriate sgRNAs each.

Treatment	**Protoplast genotype**	**Transformation constructs**
1 (Negative control)	WT	None
2 (*ubi1* sgRNA negative control)	WT	*ubi1* sgRNAs 1-4
3 (*Ppd-D1* sgRNA negative control)	WT	*Ppd-D1* sgRNAs 1-5
4 (SunTag negative control)	B3598.6	None
5 (*ubi1* sgRNA experiment)	B3598.6	*ubi1* sgRNAs 1-4
6 (*Ppd-D1* sgRNA experiment)	B3598.6	*Ppd-D1* sgRNAs 1-5

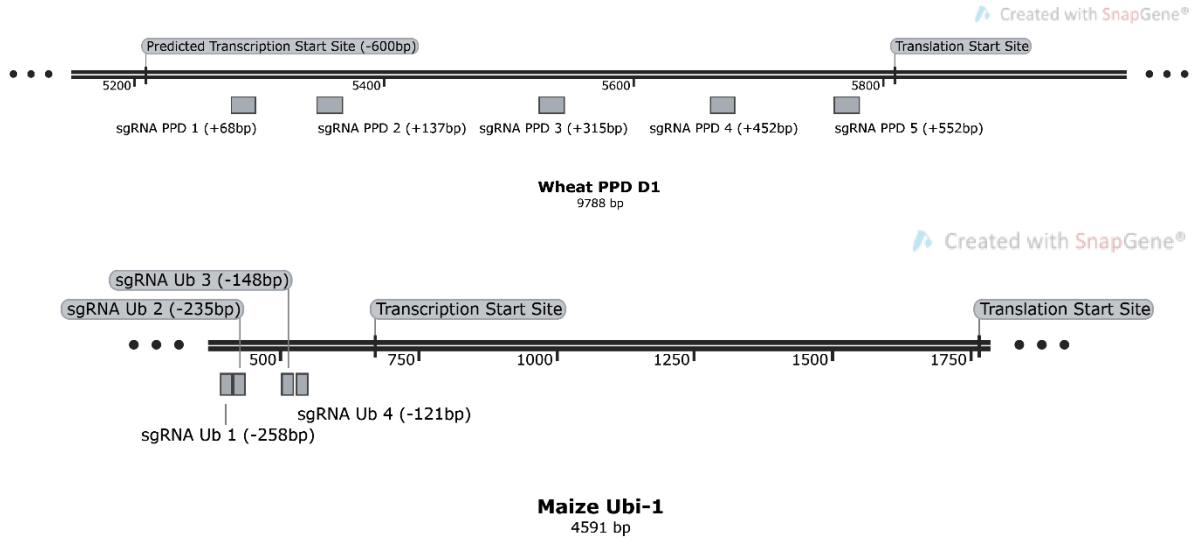

Figure 2.2: The locations of the sgRNAs designed to target the wheat *Ppd-D1* and maize *ubi1* promoters. The transcription and translation start sites are marked. The location in base pairs is in relation to the transcription start site (0).

Transformed protoplasts were incubated for 41 h at 23 °C, before total RNA was isolated using the protocol outlined in section 2.4.1, with the modification that the RNA was eluted from the QIAGEN clean-up column using 50 µl of RNase-free water.

The RNA samples were quantified and quality assessed using a Nanodrop 1000 spectrophotometer as outlined in section 2.4.1, as well as an Agilent Technologies, Inc. (Santa Clara, CA, USA) TapeStation RNA screentape (carried out by Dr Christy Waterfall (University of Bristol Genomics Facility; Bristol, UK)).

2.6.2 Library Preparation

Library preparations were carried out by Dr Christy Waterfall (University of Bristol Genomics Facility; Bristol, UK) using the Illumina, Inc. (San Diego, CA, USA) TruSeq® Stranded mRNA library prep kit according to the manufacturer's instructions. 270 ng of total RNA was used per library preparation. The cDNA libraries underwent quality control using an Agilent Technologies, Inc. TapeStation D1000 ScreenTape® assay.

2.6.3 Sequencing

Sequencing was carried out by Dr Christy Waterfall (University of Bristol Genomics Facility; Bristol, UK). An Illumina, Inc. NextSeq500 system was used with the high output version 2.5 paired-end 2x75 bp kit according to the manufacturer's instructions. NextSeq Control software v2.2.0 (Illumina, 2018) was used for run setup and monitoring. Primary analysis was conducted using Real Time Analysis v2.4.11 software (Illumina, 2015). 96 FASTQ files were generated using BaseSpace (Illumina, 2019).

2.6.4 Data Analysis

Data underwent initial quality-control using FastQC (Andrews, 2018). No trimming of the reads was deemed necessary. The FASTQ files were concatenated by Dr Alex Paterson (University of Bristol Genomics Facility; Bristol, UK) to create forward and reverse FASTQ files for each sample. These merged, un-trimmed FASTQ files were aligned to the IWGSC RefSeq v1.0 full genome sequence (IWGSC *et al*., 2018) using STAR RNA-seq aligner software (Dobin *et al*., 2013) by Dr Alex Paterson (University of Bristol Genomics Facility; Bristol, UK). Parameters used for alignments which were not set to the default value are listed in the section A.5. featureCounts (Liao *et al*., 2014) was used to assign these aligned reads to genomic features using the RefSeq v1.1 annotation (Unité de Recherche Génomique Info, 2019) by Dr Alex Paterson (University of Bristol Genomics Facility; Bristol, UK). Parameters used for alignments which were not set to the default value are listed in section A.6. DESeq2 (Love *et al*., 2014) was then used (with all parameters set to default values) to identify differential expression between samples using a negative binomial generalised linear model.

CHAPTER 3: VERIFYING THE SYSTEM

To establish that a SunTag system can induce specific gene expression in bread wheat, the presence of all SunTag system components must be verified. Secondly, the system must be put into practice and its impact on gene expression be assessed.

This results chapter lays out the steps taken to verify that two of the essential elements of the SunTag CRISPRa system (dCas9-SunTag$_{x10}$ and VP64-ScFv) can be transcribed, translated, and interact as predicted *in vitro* and *in planta*. This would satisfy the first and second aims of this project as set out in section 1.7. Theoretically, only the addition of sgRNAs (which have been previously expressed in wheat (Shan *et al*., 2014)) would be required to induce increased gene expression.

3.1 TRANSFORMATION AND GENOTYPING

3.1.1 Transformation Constructs

As stated in section 2.1, the SunTag constructs published by Tanenbaum *et al.* (2014) were modified for use in wheat as they were originally designed for use in human cell lines. One major change was codon-optimisation according to codon usage bias in wheat. Several codons (groups of three bases) can encode a single amino acid; there are 61 possible codons (excluding the three stop codons) but only 20 amino acids (Behura & Severson, 2013). The genetic code is therefore described as 'degenerate' (Crick *et al*., 1961). Where several codons encode a particular amino acid, it has been shown that organisms use synonymous codons at different frequencies. Codon usage can significantly affect gene expression, so when expressing an exogenous construct, it is important to use codons at the same frequency as the host organism (Plotkin & Kudla, 2011). For instance, when expressing interleukin-1 (IL1) and interleukin-2 (IL2) from bovine papillomavirus type 1 (BPV1) in Cos-1 cell cultures originating from *Chlorocebus aethiops* (African green monkey), codon-optimised constructs have been shown to produce 1000-fold more protein than wild-type constructs (Zhou *et al*., 1999). This reinforces the importance of codon optimisation when attempting to express constructs such as those used in the SunTag system. The maize *ubiquitin1* (*ubi1*) promoter (a constitutive promoter suitable for use in monocots) was used to drive expression of the SunTag constructs instead of the simian vacuolating virus 40 (SV40) promoter (commonly used in mammalian cell lines). This alteration was necessary as promoters do not function in the same manner in all organisms, possibly due to differing endogenous promoter architectures (Yamamoto *et al*., 2007), or varying transcription factor availabilities (Biłas *et al*., 2016). *ubi1* is a well-characterised promoter with generally strong expression (Christensen *et al*., 1992; Rooke *et al*., 2000) and is, therefore, suitable to drive expression of the SunTag constructs.

3.1.2 Transient Expression *in vitro*

In wheat protoplasts, transient expression of the modified SunTag constructs was accomplished via PEG-mediated plasmid transformation. Using a GFP reporter plasmid, a mean transformation rate of 60% was achieved.

dCas9-SunTag$_{x10}$ (Figure A.2) and VP64-ScFv (Figure A.4) plasmids were used to transform protoplasts both separately and in co-transformations. As both constructs were GFP tagged, a transformation rate was able to be estimated via fluorescent microscopy (Figures 3.1 and A.5).

Mean transformation rates for the SunTag constructs were lower than that of the GFP reporter control; 5.25% (number of studies (n) = 4) for dCas9-SunTag$_{x10}$, 18.75% (n=4) for VP64-ScFv, and 26.25% (n=4) for co-transformations. In co-transformation trials, it was not possible to determine whether protoplasts were transformed with only one or both plasmids as both carried GFP as a reporter gene. The reasons behind the lower transformation rate observed when using the SunTag plasmids compared to the GFP control plasmid are unclear; there are several possibilities.

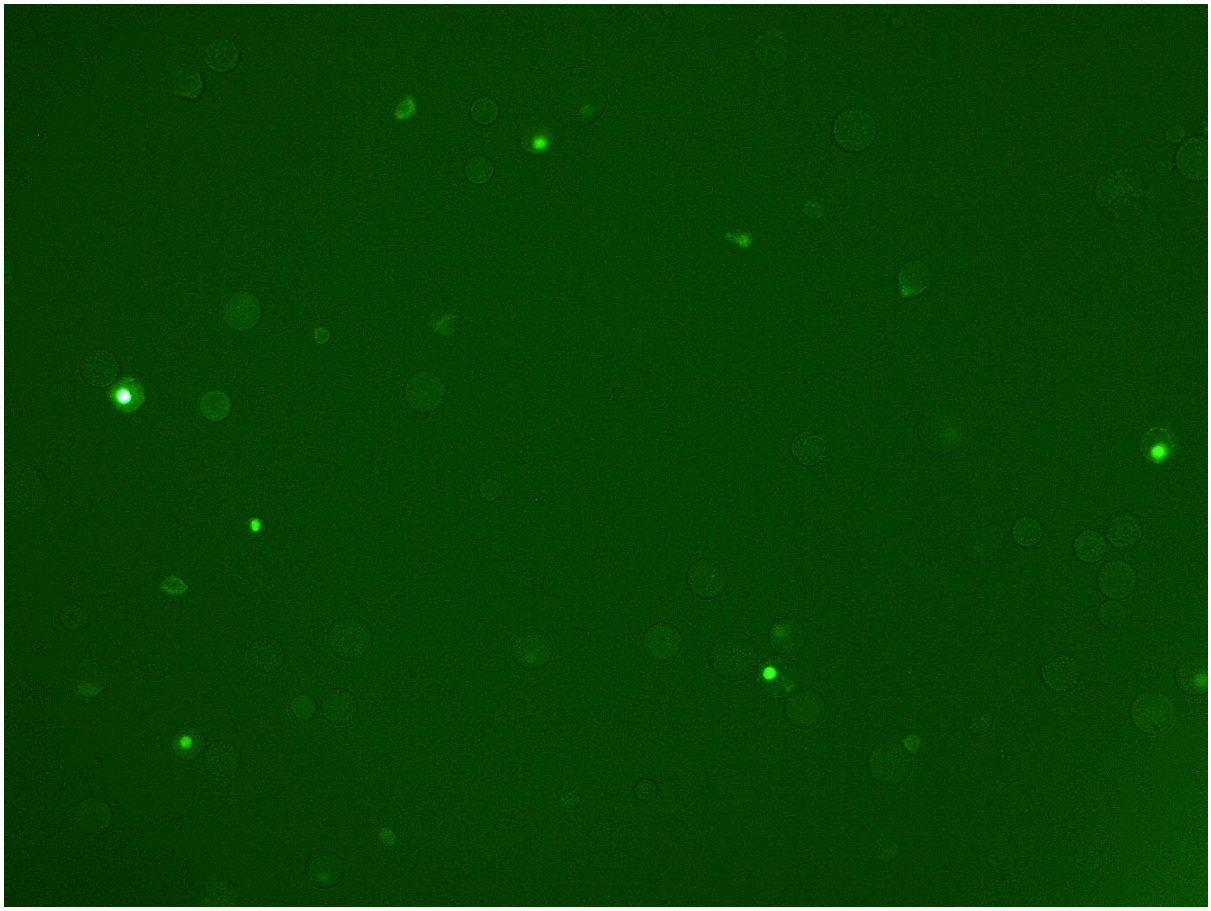

Figure 3.1: Image taken of protoplasts co-transformed with VP64-ScFv and dCas9-SunTag$_{x10}$ plasmids. The protoplasts were incubated for approximately 30 h. Excitation: 470 nm; excitation intensity: 20%; magnification: 10x; filter cube: GFP.

If the transformation efficiency for the SunTag constructs was low, as indicated by fluorescent microscopy, it was unlikely to be due to the transformation technique used as the 60% mean transformation rate achieved using the GFP reporter plasmid is similar to that achieved by Shan *et al.* (2014). The dCas9-SunTag$_{x10}$ plasmid, at over 10 kb, is much larger than either the VP64-ScFv or GFP plasmids which are both 6 kb. This may mean that the plasmid may be less likely to enter the protoplasts. It has been shown in *E. coli* that plasmid size is negatively correlated with transformation rate (Chan *et al*., 2002; Hanahan, 1983; Rehman *et al*., 2016). This may also be the case in wheat protoplasts. However, this would not explain the lower transformation rate observed for the VP64-ScFv plasmid, as it is just 25 bp larger than the GFP reporter plasmid.

There is also the possibility that using the number of observable GFP signals as a proxy for transformation rate is inaccurate. The transformation rate may have been higher than estimated, however, despite cells containing construct DNA, some of the GFP tags may not have been present or have been too weak to have been observed. That is, transformed protoplasts may have shown low or undetectable levels of fluorescence. Whether a GFP signal is observed is partially dependent on the sensitivity of the detection system, including the microscope, camera and the eyes of the researcher. A more sensitive technique could be trialled, such as confocal microscopy; this technique generally generates images with a lower level of background (Fellers & Davidson, 2019) making any GFP signal easier to identify. Additionally, the excitation and capture parameters used in this study were not optimal. For instance, the excitation peak of sfGFP is 485 nm, while the wavelength used was 470 nm. Therefore, weaker GFP signals may have not been identified.

However, there may have also been issues with the fluorescent protein pre-translation, during translation, or post-translation which led to a low concentration of GFP. Pre-translational issues would involve construct DNA or mRNA. The promoter may not express the constructs constitutively as expected. This is unlikely, as the *ubi1* promoter has been used previously to successfully drive exogenous constructs in wheat (Wang *et al*., 2014b). However, there has been some evidence that *ubi1* does not always act as a constitutive promoter; it is most active in "young, metabolically active tissues and in pollen grains" (Rooke *et al*., 2000). The protoplasts analysed in this study were isolated from 9-12 d old shoots. As one would assume that these tissues are metabolically active and so show constitutive *ubi1* expression, it is unlikely that failure of the promoter accounts for low GFP expression. *ubi1* activity could be increased through heat shock at the seedling stage, a strategy which has been shown to increase *ubi1* activity in maize (Christensen *et al*., 1992). It has been previously observed that using the same promoter to express multiple transgenes can result in transcriptional gene silencing due to DNA methylation (Peremarti *et al*., 2010). Therefore, the system could be altered so different promoters are used for each construct, so eliminating this possibility.

Construct mRNA may not mature correctly in wheat, or it may experience a high level of degradation. Both incomplete mRNA maturation and mRNA degradation are processes that are used by host cells during the regulation of gene expression, and either may have been triggered by a specific motif in the SunTag constructs (Tourrière *et al*., 2002). There are no known reasons why this would occur, and this uncertainty is why the presence of correctly processed SunTag construct mRNA was investigated as described in section 3.2.

Translational issues might account for the low number of GFP signals observed. Early termination can occur when release factors (proteins which govern translation termination) inaccurately identify sense codons (which encode an amino acid) as stop codons (Freistroffer *et al*., 2000). Early termination of protein translation would produce a truncated protein, almost certainly leading to a lack of GFP expression in the dCas9-SunTag$_{x10}$ protein as the sfGFP is located at the C-terminus of the construct. However, there is a small possibility that translation was reinitiated later in the mRNA, therefore producing a GFP signal where the SunTag construct is not being completely synthesised. Termination and reinitiation of mRNA molecules has been shown to occur in mammalian expression vectors (Peabody & Berg, 1986). The sfGFP motif is near the N-terminus of the VP64-ScFv protein, so any early translation termination would not necessarily interfere with the GFP signal. This uncertainty necessitated an investigation into the presence of the SunTag proteins in transformed protoplasts and plants; this is explored in section 3.3.

Finally, the SunTag proteins may be unstable and become degraded, although there is no known reason why the VP64-ScFv protein would be unstable in wheat. Other ScFv proteins have been shown to express in wheat (Stöger *et al*., 2000), however, there are no examples of either VP16, or its tetramer VP64, being successfully expressed in wheat. Therefore, it is not known whether this protein is stable in wheat cells. Constructs containing dCas9 have been expressed in wheat cells (Zong *et al*., 2017), however, Tanenbaum *et al.* (2014) reported issues with dCas9-SunTag$_{x10}$ protein stability during the initial development of the SunTag system. The group made efforts to increase the stability of the protein by decreasing the number of hydrophobic residues and increasing α-helical propensity (Tanenbaum *et al*., 2014). However, this may still pose an issue and lead to a low SunTag protein concentration, and therefore a difficult to detect GFP signal. It is important to note that the dCas9-SunTag$_{x10}$ construct produces separate dCas9-SunTag$_{x10}$ and GFP proteins. A porcine teschovirus-1 2A (P2A) motif separating dCas9-SunTag$_{x10}$ and GFP induces a ribosome skipping mechanism whereby a glycyl-prolyl bond within the P2A amino acid sequence is 'skipped', so producing separate proteins (Donnelly *et al*., 2001). Therefore, dCas9-SunTag$_{x10}$ protein instability wouldn't necessarily affect GFP expression.

Although these potential issues, before, during and after the translation process may mean that the transformation rate is higher than estimated, problems such as aberrant transcription or protein instability may result in a less efficient, or even non-functional CRISPRa system. If the number of GFP signals is an accurate proxy, and the expression level of the two SunTag constructs was low, this may not pose an issue as only a few dCas9-SunTag$_{x10}$ molecules are needed for the system to function (Tanenbaum *et al*., 2014). Furthermore, the lead author of the paper suggested that a low level of SunTag construct expression is, in fact, optimal (personal communication). Therefore, the study was continued using these constructs despite the apparently low transformation rates, and further steps were taken to verify the presence of SunTag construct mRNA and protein.

GFP signals were occasionally observed outside of the nucleus (Figure 3.2). This suggests that the protein may have been aggregating and was not being transported to the nucleus. The phenomenon was observed by Tanenbaum *et al*. (2014) during the development of the SunTag ScFv; the antibody was modified to reduce this phenomenon. These trials were conducted in human cells so further alterations may need to be made to optimise the system for wheat to eliminate the issue. Aggregation was not deemed to be an important issue, as it was also observed by Dr Mark Winfield (University of Bristol; Bristol, UK) in protoplasts in successful Cas9 editing trials (personal communication), and the phenomenon was observed in <0.5% of cells. Therefore, it was determined that the study should be continued despite this observation. In future, the structure of

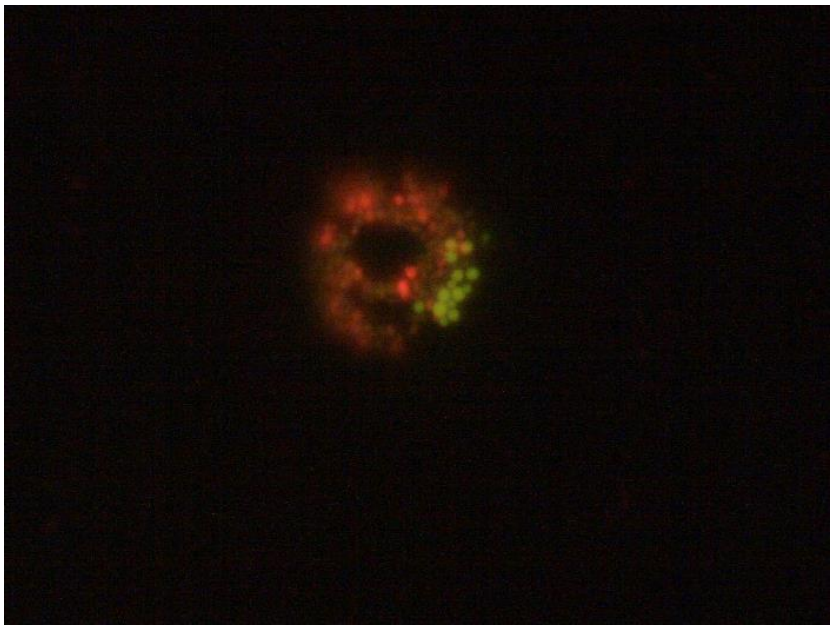

Figure 3.2: Image taken of a protoplast transformed with the dCas9-SunTag$_{x10}$ plasmid. The red signal in the cytoplasm is chloroplast autofluorescence. The protoplasts were incubated for approximately 46 h. Excitation: 470 nm; excitation intensity: 20%; magnification: 40x; filter cube: BGR.

the protein could be altered so that there is a lower likelihood of tertiary structures being formed. A stronger nuclear localisation sequence, or a higher number of nuclear localisation motifs, may also decrease the possibility of these accumulations.

3.1.3 Stable Expression *in planta*

Lines that stably express the SunTag constructs were generated by simultaneous particle bombardment using the same constructs as used in protoplasts, with the slight adjustment that dCas9-SunTag$_{x10}$ was tagged with yellow fluorescent protein (YFP) rather than GFP.

Genotyping of whole transformed plants was conducted by amplification of sections of the constructs from whole genomic DNA extracted from young plants. Primers specific to an adenosine diphosphate-ribosylation factor (ARF) gene (Kobayashi-Uehara *et al*., 2001) were used as a positive control. Genotyping was carried out on nine B3598 plants that germinated from 12 seeds planted. It was established that two plants were positive for both constructs; from this limited number of individuals, it appeared as though the constructs were co-segregating at this generation (Figure 3.3). The two positive B3598 plants (B3598.6 and B3598.9) and a negative plant (B3598.1) were carried forward for further study. Seed was collected from B3598.1, B3598.6 and B3598.9, and the thousand-grain weight was calculated to be 40.99 g, 31.87 g, and 31.49 g respectively. T3 progeny of B3598.6 and B3598.9, as well as T3 individuals of the E3 line, were also genotyped and 48.72% were positive for the SunTag constructs (n=39). This process established that the constructs necessary for the SunTag system had been successfully integrated into the genome.

The number of genotyped plants that proved positive was lower than expected based on Mendelian ratios (Mendel, 1865); only 43.75% contained both SunTag constructs (n = 48), which appeared from the relatively small sample size to be co-segregating. The laws of Mendelian genetics would predict that 75% the progeny of a hemizygous parent to contain co-segregating constructs (Mendel, 1865). There are several possible reasons for this unexpected observation. The presence of the SunTag constructs may decrease fitness, so skewing the genotype of the seeds towards null individuals. Few studies have been conducted on the fitness costs of biolistic transformation, and those that have been carried out are based upon herbicide- or antibiotic-resistant transgenic plants (Guadagnuolo *et al*., 2006; Purrington & Bergelson, 1997). The evidence presented in these studies is somewhat conflicting, with a fitness cost being incurred for some transgenes and not others (Purrington & Bergelson, 1997; Guadagnuolo *et al*., 2006). Across five measures of fitness and fecundity, no statistically significant differences were observed between transgenic and non-transgenic progeny of a selfed maize plant hemizygous for glyphosate resistance (Guadagnuolo *et al*., 2006). Purrington & Bergelson (1997) found that there was a fitness cost in *A. thaliana* individuals containing a herbicide

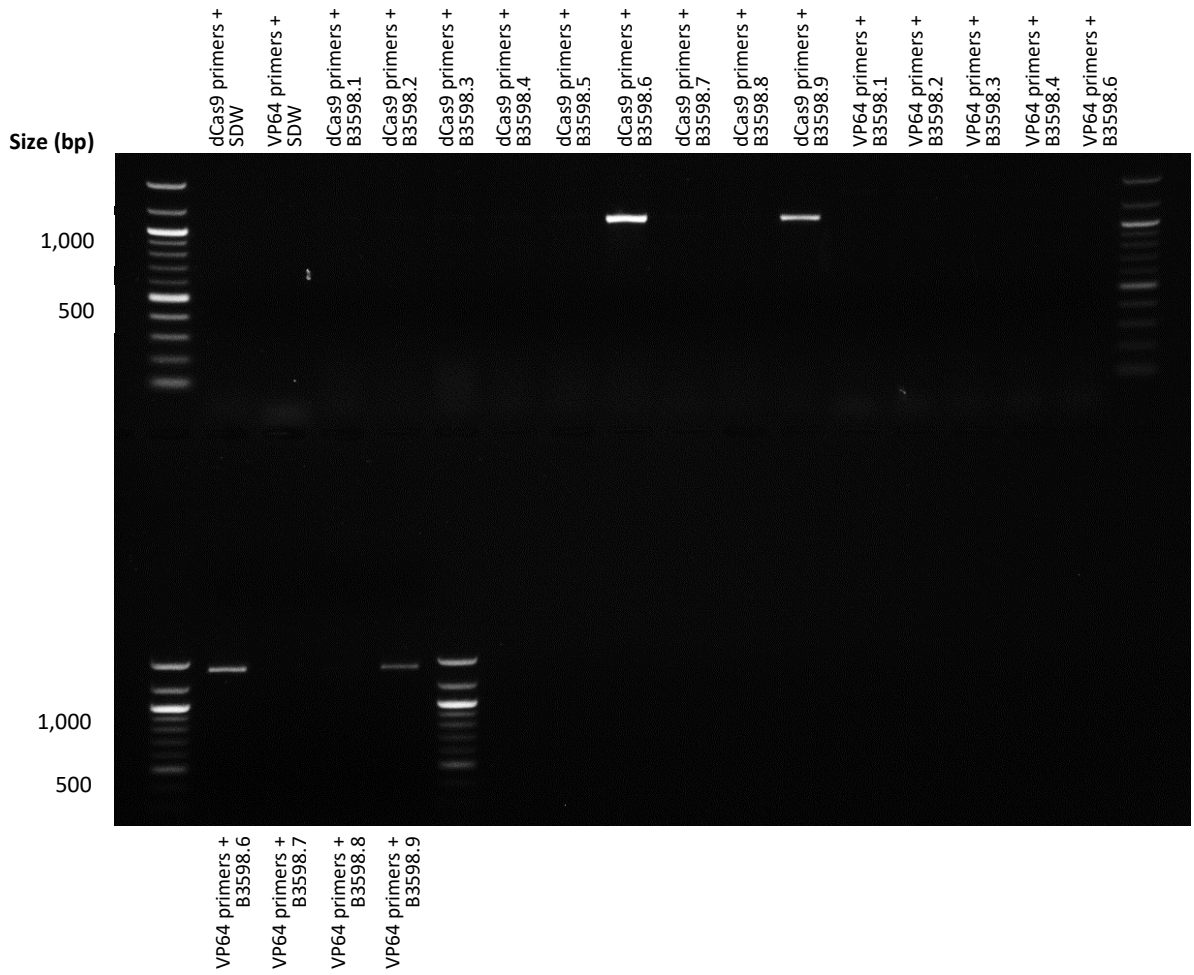

Figure 3.3: Gel electrophoresis (1% w/v agarose gel, 100 V, 2 h) of B3598 PCR products. The two primer sets used were dCas9 (Ub F & dCas9 R) and VP64 (Ub F & VP64-ScFv R). A New England Biolabs Inc. (Waltham, MA, USA) QuickLoad 100bp DNA ladder was used

resistance gene, but not in those containing an antibiotic resistance gene or an empty plasmid. This suggests that the presence of a transgene itself is not detrimental due to resources being directed away from host growth and metabolism, commonly known as 'metabolic drain' (Glick, 1995). The antibiotic resistance plasmid found to not impart a fitness cost was 0.00872% of the size of the *A. thaliana* genome; the SunTag constructs inserted into the stably transformed wheat plants are, combined, just 0.000102% of the wheat genome. Theoretically, therefore, the SunTag constructs are too small to place a metabolic burden on the plants.

The dCas9-SunTag$_{x10}$ or VP64-ScFv proteins may interact with an endogenous gene in a deleterious manner, such as through competitive inhibition. This is not an unlikely scenario, as the constructs contain some common epitopes as well as motifs that are also present in the wheat genome, such as the H2B nuclear localisation signal. Therefore, any interactions that involve endogenous copies of H2B

could feasibly also occur with the copy on the dCas9-SunTag$_{x10}$ protein. To investigate this further, a deeper analysis of the other proteins pulled out by the co-IP experiments described in section 3.4 could be conducted to elucidate any possible unforeseen interactions. A further co-IP and LC-MS analysis of its products should be conducted using antibodies specific to the VP64-ScFv protein to discover any possible detrimental interactions between VP64-ScFv and endogenous proteins. This result, alongside the low transformation rate observed in protoplasts, suggests one or both SunTag constructs may be deleterious in wheat. In all protoplasting trials, some protoplasts were observed to have died. Whether these were transformed or wild type (WT) cells could not be determined as any GFP could have degraded, and cell death causes autofluorescence (Grønlund *et al*., 2012). Therefore, it is unknown whether transformed cells had a higher death rate than WT protoplasts. However, it is possible to sort protoplasts using fluorescence-activated cell sorting (FACS), which has previously been used to separate transformed, non-transformed, and dead or dying *A. thaliana* protoplasts (Grønlund *et al*., 2012). The percentage of each type of cell transformed could then be qualitatively estimated via PCR to determine whether there is a substantial difference between the transformation rate in protoplasts that have and have not survived, although this method would not be likely to detect more subtle differences in transformation rates.

The lower thousand seed weight for plants containing the SunTag constructs is additional evidence for the constructs being detrimental to plant fitness. However, the sample sizes were not large enough to determine statistical significance. Further work, such as the FACS study described previously, is needed to clarify whether the constructs are deleterious or not.

A low transformation rate would not pose a serious issue during stable transformation, as the constructs only need to be incorporated into the genome once to eventually create a stably expressing non-segregating line. This issue may, however, necessitate the bombardment of a larger number of calli to successfully obtain transformants.

3.2 TRANSCRIPTION

The presence of dCas9-SunTag$_{x10}$ and VP64-ScFv mRNA, both *in vitro* and *in planta*, was verified by RT-PCR conducted on total RNA extracted from both protoplasts and plants B3598.6 and B3598.9 (Figures 3.4 and 3.5). The sizes of the bands correspond to those of correctly spliced mRNA. The sizes of the cDNA fragments are smaller than those of the full DNA sequences because the *ubi1* promoter intron is spliced out during mRNA maturation. Therefore, the smaller fragment sizes observed in Figures 3.4 and 3.5 are an indication that the mRNA was being processed correctly.

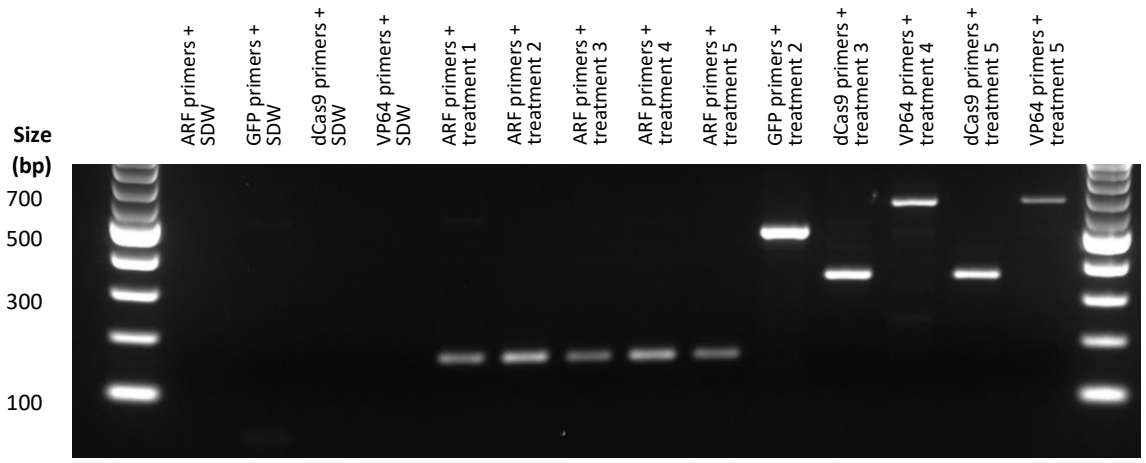

Figure 3.4: Gel electrophoresis (2% w/v agarose gel, 100 v, 2 h 30 min) of protoplast RT-PCR products. There were five sets of protoplasts to which different transformation treatments were applied: 1) negative control; 2) GFP reporter plasmid; 3) dCas9-SunTag$_{x10}$ plasmid: 4) VP64-ScFv plasmid; 5) co-transformation with the dCas9-SunTag$_{x10}$ and VP64-ScFv plasmids. The four primer sets used were ARF (ARF (F) & ARF (R)), GFP (Ub (F) & H2B (R)), dCas9 (Ub (F) & dCas9 (R)), and VP64 (Ub (F) and VP64-ScFv (R)). A New England Biolabs Inc. (Waltham, MA, USA) QuickLoad 1 kb+ DNA ladder was used.

It also suggests that there was no DNA contamination, which would be expected to consist of the larger, unspliced fragments. To explicitly show that these products were amplified from SunTag construct RNA, the cDNA was sequenced and aligned to the expected sequence. All products sequenced were as expected, including correct splicing of the *ubi1* promoter intron.

The smearing in Figure 3.5 might indicate an issue with the RT-PCR. One explanation may be incorrect loading. The quantity of RNA used as a template for the RT-PCRs was 500ng, within the range recommended by the RT-PCR reaction mix manufacturer. The concentration of each RNA sample, and therefore the volume of each sample to be added to each reaction, was quantified using a Nanodrop spectrophotometer. This method does not distinguish between RNA and DNA, however, as the RNA samples were DNase treated it is unlikely there was any DNA contamination which would make the concentration readings unreliable. The number of cycles used may have been too high, producing a higher than expected concentration of product. However, in preliminary studies, a lower number of cycles did not produce any visible bands during gel electrophoresis. It is unlikely to be the result of DNA crossover or contamination as the negative control lanes are free from any signal. It may be due to a non-optimal magnesium ion concentration, which can lead to a high level of background. This is somewhat unlikely, as the reaction buffer was used according to the manufacturer's

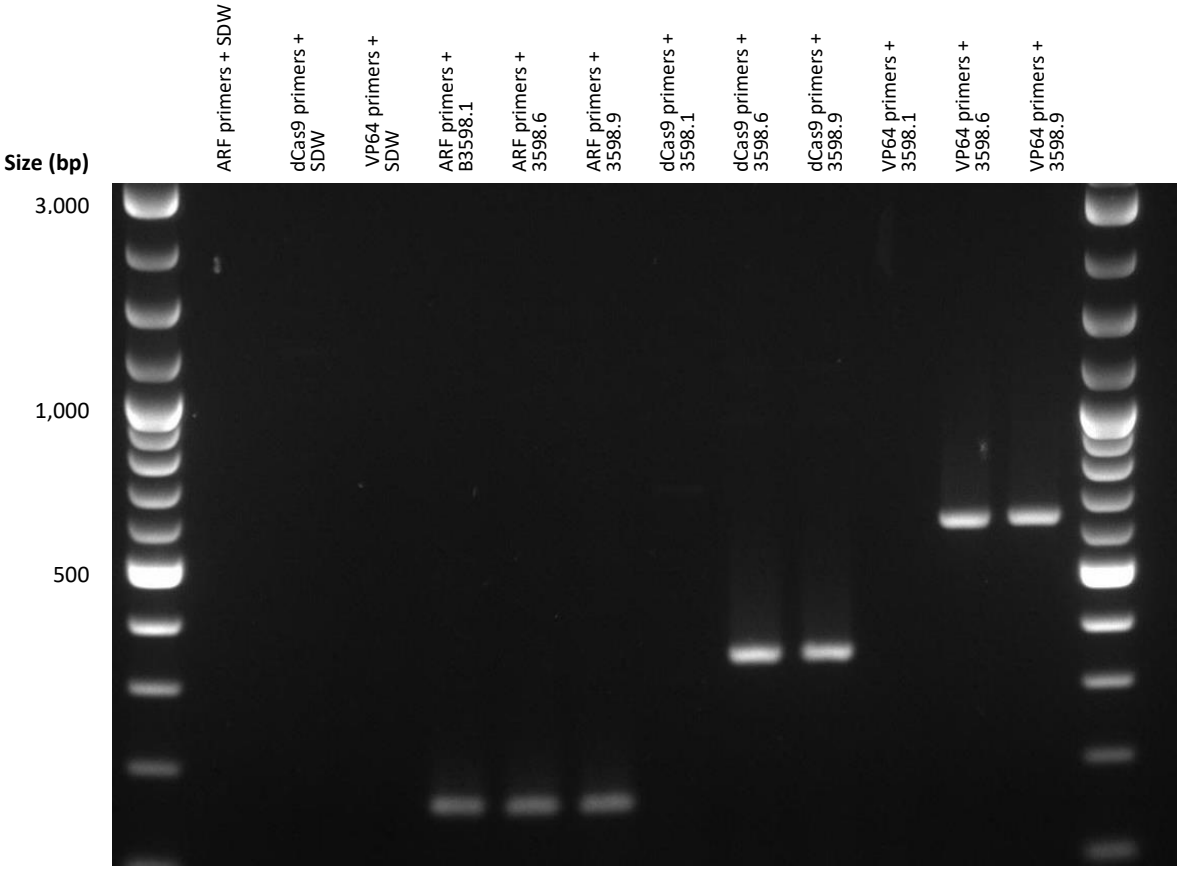

Figure 3.5: Gel electrophoresis (2% w/v agarose gel, 80 v, 4 h) of B3598 RT-PCR products. The three primer sets used were ARF (ARF (F) & ARF (R)), dCas9 (Ub (F) & dCas9 (R)), and VP64 (Ub (F) and VP64-ScFv (R)). A New England Biolabs Inc. (Waltham, MA, USA) QuickLoad 1 kb+ DNA ladder was used.

instructions, which eliminates the need for magnesium ion concentration optimisation. However, this could be an unusual case where a more 'extreme' magnesium ion concentration is necessary. 0.015 µM of each primer was used, which is far lower than the 0.6 µM which QIAGEN recommends. This decision was based on previous studies carried out within the Edwards' group; however, a higher concentration could be trialled.

Despite these possible issues, the fact that the smears are not seen in negative controls but are also seen in positive controls suggests this is an issue with procedure rather than the starting material. In future, given more time, further optimisation of the protocol should be carried out.

3.3 TRANSLATION

To ensure that construct mRNA was subsequently being translated into dCas9-SunTag$_{x10}$ and VP64-ScFv proteins, Western blots were carried out on total protein extracted from stably transformed plants. The antibody used was 'Cas9 antibody 3', the details of which are given in

Table A.1. Two other antibodies were also produced at the same time against two different peptides, however, in trials carried out by Professor Keith Edwards (University of Bristol; Bristol, UK) these were found to be less sensitive and less specific than antibody 3 (personal communication). A commercial Cas9 nuclease was used as a positive control in the form of serial dilutions in all blots, however, in trials, it was noted that this did not appear to be running at the expected 159 kDa size for the protein (Figure 3.6). To see whether this was as a result of the protein being degraded (as it was running smaller than expected) or the size standards not running at the expected size, a different Cas9 nuclease (synthesised by Dr Lucy Hyde (University of Bristol; Bristol, UK)) was run on the same blot as

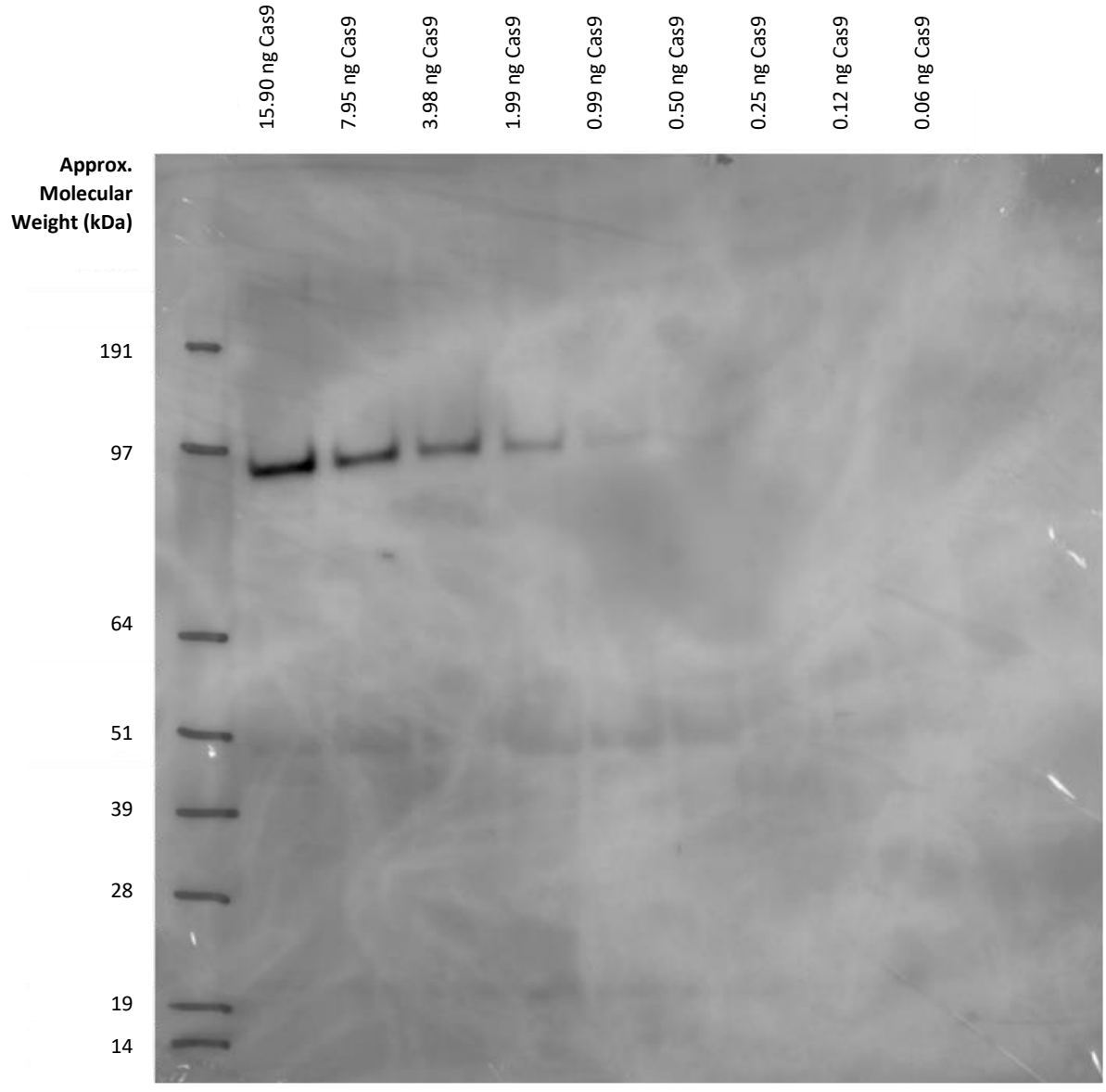

Figure 3.6: Western blot of New England Biolabs Inc. (Waltham, MA, USA) Cas9 nuclease. The mass of Cas9 nuclease stated for each lane is suspended in WT wheat cv. Cadenza total protein extract. Primary antibody: Cas9 antibody 3 (1 µg/ml).

a serial dilution (Figure 3.7). In this trial, both Cas9 proteins migrated the same distance on the gel, which corresponded to the incorrect size according to the SeeBlue™ Plus 2 pre-stained protein standards. Therefore, it was concluded that the size standard could not be relied upon for protein size determination. It was, however, still included in Western blot experiments as an indication of how far the gel had run, and therefore for how long the voltage should be applied.

No signal was detected for total protein extracted from plant B3598.6 (Figure 3.8). This suggests that either the level of SunTag construct expression is low, or the proteins are not being expressed at all. However, there is also the possibility that the protein was not being extracted using the Sigma-Aldrich

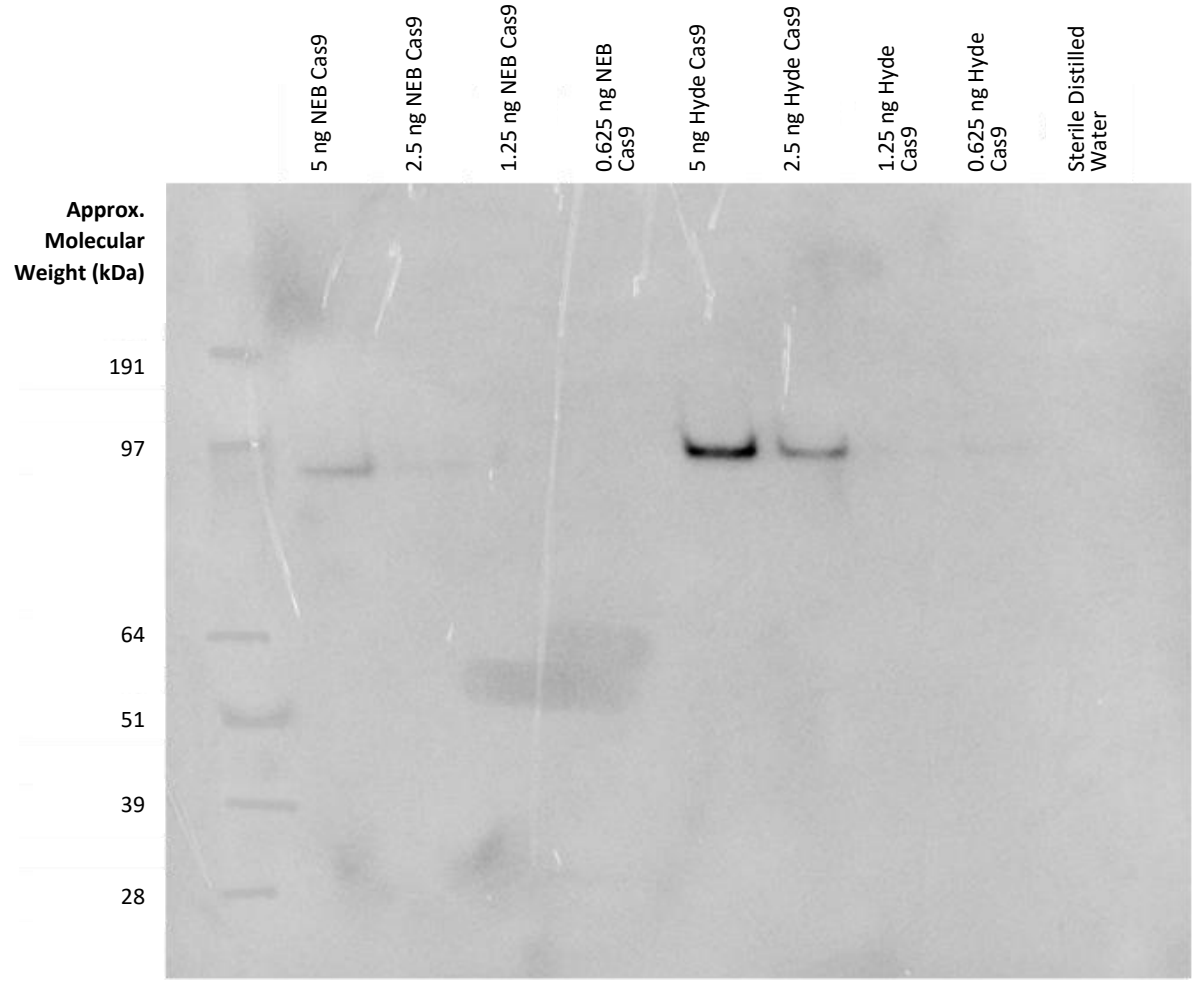

Figure 3.7: Western blot of New England Biolabs Inc. (Waltham, MA, USA) Cas9 nuclease and Cas9 protein synthesised by Dr Lucy Hyde. Primary antibody: Cas9 antibody 3 (1 µg/ml).

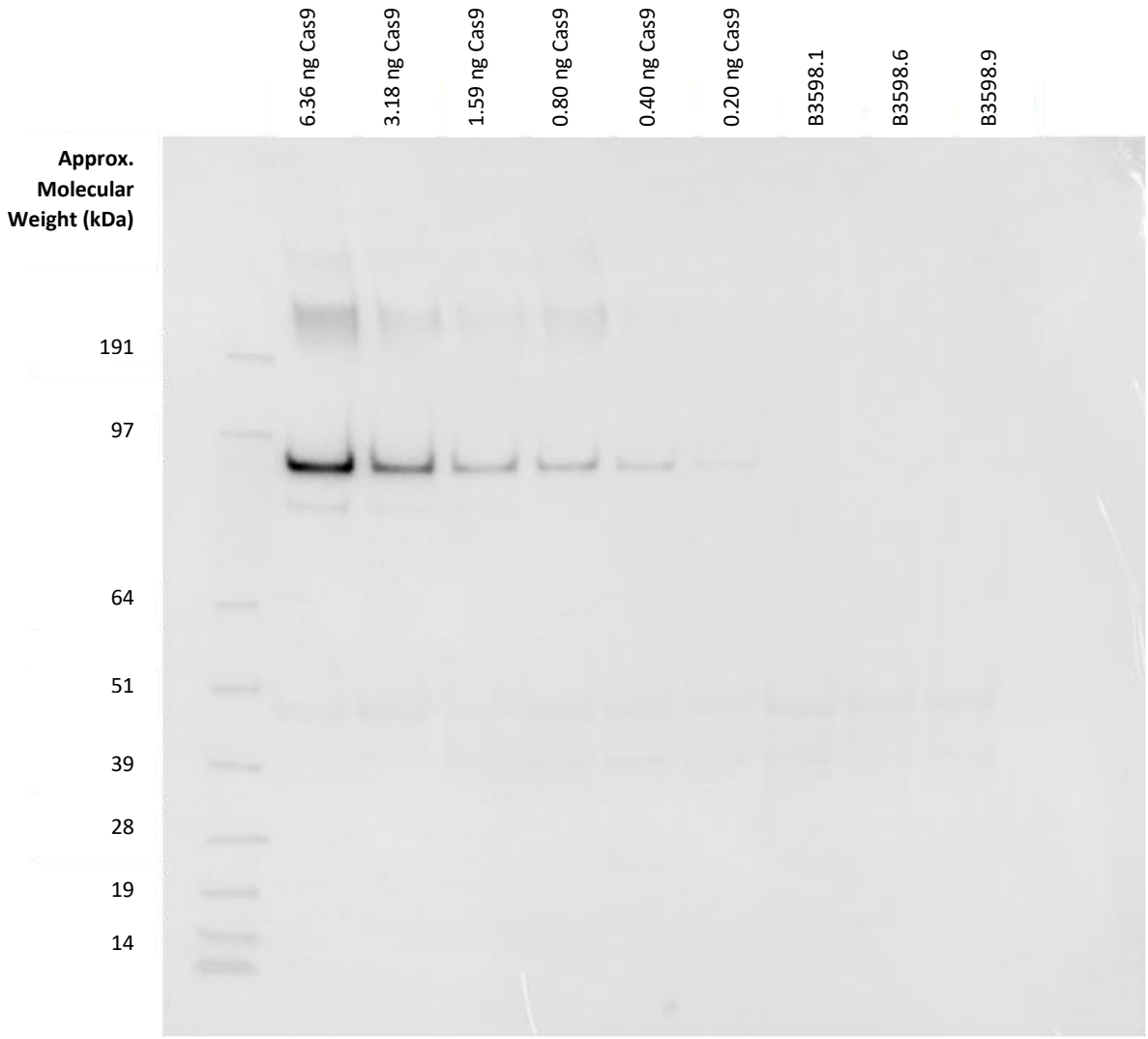

Figure 3.8: Western blot of B3598.1, 6 and 9 total protein. New England Biolabs Inc. (Waltham, MA, USA) Cas9 nuclease is a positive control. The mass of Cas9 nuclease stated for each lane is suspended in WT wheat cv. Cadenza total protein extract. Primary antibody: Cas9 antibody 3 (1 µg/ml).

protein extraction kit (as described in section 2.5.1). This may be due to the dCas9-SunTag$_{x10}$ protein not being soluble in the lysis buffer. Therefore, two other techniques were tested (as described in sections 2.5.2 and 2.5.6), however, a negative result was also obtained for these samples. The sensitivity achieved for these Western blots was around 2ng of protein per sample. Therefore other, more sensitive techniques were necessary to detect any possible low level of construct expression.

Samples were sent to Dr Kate Heesom at the University of Bristol proteomics facility for liquid chromatography-mass spectrometry (LC-MS) analysis. Cell lysate extracted from plants B3598.6 and B3598.9 was used in the initial study. No peptides matching the VP64-ScFv protein were detected. The

only peptides detected which matched the dCas9-SunTag$_{x10}$ protein amino acid sequence were part of the H2B sequence included in the construct as a nuclear localisation signal. The peptides detected are also present in the endogenous copy of H2B in wheat, and therefore the source of these peptides could not be determined. It was concluded that the peptides were unlikely to be from the dCas9-SunTag$_{x10}$ protein due to the absence of other peptides from other, unique regions of the construct. Therefore, the most likely explanations are that either the SunTag constructs were not present, or were being expressed at a level too low to be detected by LC-MS.

3.4 PROTEIN INTERACTION

To both enrich the protein sample for LC-MS analysis and to determine whether the VP64-ScFv and dCas9-SunTag$_{x10}$ proteins interact *in planta*, a co-immunoprecipitation (co-IP) reaction was conducted on a leaf sample from plant B3598.6. The coupling efficiency of the antibodies to the agarose beads for all three positive reactions was over 63%. The co-IP products were run on a gel and blotted using Cas9 antibody 3 (Figure 3.9). Again, no definitive signal was detected. Some bands were visible for the three positive co-IP reactions which were not present for the negative control reactions. This suggests that these proteins bound to all three Cas9 antibodies and so the detected protein has some sequence homology to the Cas9 nuclease. There is the possibility that it is the dCas9-SunTag$_{x10}$ protein although its position on the gel indicates that it was a great deal smaller than the Cas9 control protein which is 159 kDa. The expected size of the dCas9-SunTag$_{x10}$ protein is 198.16 kDa, so the signal observed in Figure 3.9 is highly unlikely to be the full dCas9-SunTag$_{x10}$ protein. The smallest dCas9-SunTag$_{x10}$ protein fragment which contains all three antibody epitopes is 126.88 kDa, so the signal may have been a degraded form of the dCas9-SunTag$_{x10}$ protein. The signal is unlikely to be the 72.81 kDa VP64-ScFv protein as the primary antibody used for the Western blot was specific to Cas9. Finally, the smaller-than-expected signal could be the result of non-specific antibody binding, although this is unlikely as the signal was not present in the negative controls.

As a result of this uncertainty, the samples were sent for LC-MS analysis, the results of which are shown in Table 3.1. Peptides uniquely matching the VP64-ScFv protein amino acid sequence were detected in all three co-IP products, and were not present in either negative control (co-IPs 4 and 5). However, once again the only peptides detected relating to the dCas9-SunTag$_{x10}$ construct were within the H2B motif also present in wild type wheat; the source of these peptides could not be determined.

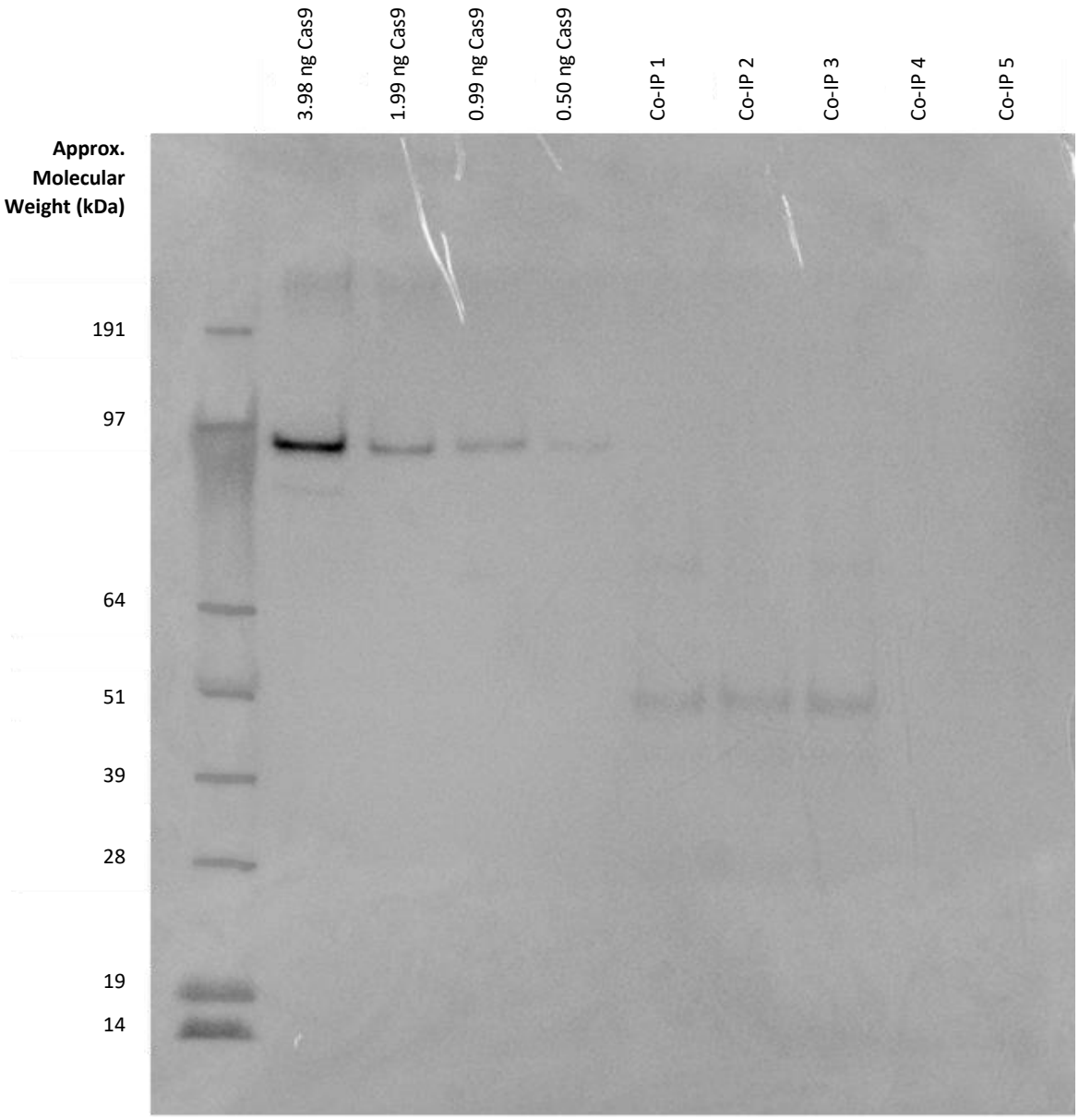

Figure 3.9: Western blot of co-immunoprecipitation products. New England Biolabs Inc. (Waltham, MA, USA) Cas9 nuclease is a positive control. Co-IP samples 1, 2, and 3 were carried out using Cas9 antibodies 1, 2, and 3 respectively bound to agarose beads. Co-IP 4 and 5 were negative controls. Primary antibody: Cas9 antibody 3 (1 µg/ml).

Although the presence of the dCas9-SunTag$_{x10}$ protein was never shown, through either Western blots or LC-MS, the results of the co-IP LC-MS run strongly suggest it is present but at levels too low to be detected. It also indicated that the two SunTag proteins are binding to each other. This is because the antibodies bound to the agarose beads should pull out the dCas9-SunTag$_{x10}$ construct and any proteins bound to it; the presence of VP64-ScFv peptides is dependent on the presence of the dCas9-SunTag$_{x10}$ protein which it theoretically binds to. There are, however, alternative possible explanations for this result. The VP64-ScFv protein may be non-specifically binding to the Cas9 antibody directly. This is

Table 3.1: LC-MS analysis of co-IP products detailing the peptide hits aligning uniquely to the proteins of interest (VP64-ScFv and dCas9-SunTag$_{x10}$). 'Coverage' indicates the percentage of the protein sequence covered by the peptides detected. 'Peptides' indicates the number of different peptides detected that match the protein of interest. 'Unique peptides' indicates the number of peptides detected that are unique to the protein of interest. 'Area' is an indication of protein abundance, which is calculated as the average area under the three most abundant peaks for the protein of interest.

Treatment	Protein of interest	Coverage	Peptides	Unique peptides	Area
Co-IP 1	VP64-ScFv	7.42	3	3	2183724.938
Co-IP 2	VP64-ScFv	5.09	3	3	4092338.219
Co-IP 3	VP64-ScFv	9.34	3	3	3453820.135
Co-IP 4	No hits				
Co-IP 5	No hits				

unlikely as the amino acid sequences of the peptides the three Cas9 antibodies were raised against are not present in the VP64-ScFv amino acid sequence. It would additionally have to bind non-specifically to all three antibodies (as it was detected in all three co-IP reaction products). The protein may bind to the agarose beads, however, this is also unlikely as the cell lysate was pre-cleared, meaning any proteins that bound to the bead should have been discarded at this stage and should not be present in the final product.

A total of 264 peptides were detected by the LC-MS run that were pulled out by all three Cas9 antibodies, but not by the control treatments. This suggests the specificity of the co-IP experiment could be improved, possibly by using different antibodies or by increasing the number of wash steps. However, this would also be likely to decrease the yield of any target protein. This may explain the presence of the signal in the Western blot of the co-IP products (Figure 3.9); these may be non-specifically binding products. Additionally, the lower molecular weight bands in Figure 3.9 appear to be in a similar position to the background observed in Figure 3.6 for WT protein extractions.

The possibility of the dCas9-SunTag$_{x10}$ protein being unstable, a hypothesis raised in section 3.1.2, is unlikely considering the co-IP LC-MS results. The dCas9-SunTag$_{x10}$ protein would have to be present and intact for long enough for it to bind to the agarose beads along with the VP64-ScFv, remain bound during washing steps and subsequently be eluted.

The above strongly suggests that the dCas9-SunTag$_{x10}$ protein is present in transgenic tissue along with the detected VP64-ScFv protein and that the two proteins are interacting; the project was continued under this assumption. Subsequent experiments attempting to put the SunTag system into practice would, in any case, establish whether these essential elements of the system were present.

3.5 CONCLUSION

In this chapter the first and second aims of this study, as set out in section 1.7, have been mostly satisfied: to "Implement transient and stable expression of a modified SunTag system in wheat cells and plants" and "Verify that the dCas9-SunTag$_{x10}$ and VP64-ScFv SunTag constructs are present in wheat cells and plants, and are expressing the corresponding mRNA and proteins. Verify that these SunTag proteins are interacting".

The successful transformation of both protoplasts (*in vitro*) and plants (*in planta*) with the two novel SunTag constructs was shown through the observation of GFP tags, as well as PCRs. The presence of correctly spliced mRNA for both constructs was demonstrated through RT-PCR studies. Finally, evidence for the presence of the VP64-ScFv protein was obtained through co-IP and LC-MS, while the presence of the dCas9-SunTag$_{x10}$ protein was strongly suggested by the presence of the VP64-ScFv protein. An interaction between the two SunTag products was also indicated by the co-IP LC-MS result.

Although not conclusive, the evidence described above was determined to be strong enough to continue the study and put the SunTag system into practice to attempt to satisfy the third aim of the project, to "Induce specific gene expression of both inserted and endogenous genes, and quantify any changes in gene expression". A positive result from the RNA-seq experiment described in Chapter 4 would, in any case, provide further evidence for the presence and interaction of the two SunTag proteins.

CHAPTER 4: IMPLEMENTATION OF THE SUNTAG SYSTEM IN WHEAT PROTOPLASTS

To assess whether the SunTag system can induce specific gene expression in wheat, it was used to target the promoter regions of two genes and comparative gene expression analysis was conducted using RNA-seq. The first gene targeted was the exogenous maize *ubi1* promoter, which drives expression of the SunTag constructs (Figure 2.1). In theory, if the system were to successfully increase the activity of the *ubi1* promoter, a positive feedback loop would be created and the SunTag constructs would upregulate themselves. The *ubi1* promoter has a high, constitutive base level of expression (Christensen *et al*., 1992). The second target for the SunTag system was the endogenous *Photoperiod-D1* (*Ppd-D1*) gene. *Ppd-D1* is a well-characterised gene located on the D-genome which controls photoperiod sensitivity in wheat (Guo, 2010). Cadenza, the wheat variety used throughout this study, contains a photoperiod sensitive allele of *Ppd-D1* (Farré *et al*., 2016) which has a low base level of expression of approximately 4.72 transcripts per million (TPM) (Ramírez-González *et al*., 2018).

The aim of this chapter was to assess the effect on gene expression of the SunTag system with sgRNA pools targeted to the *ubi1* and *Ppd-D1* promoter regions in wheat protoplasts. The effect on the expression of both the target genes and the rest of the transcriptome was evaluated using RNA-seq, to assess the system's efficiency and off-target effects.

4.1 QUALITY CONTROL OF RNA-SEQ DATA

An RNA-seq run was conducted as described in section 2.6. In brief, total RNA was isolated from wheat protoplasts generated from WT and B3598.6 plants (the latter segregates for both SunTag constructs). Protoplasts were additionally transformed with sgRNAs targeting the *ubi1* or *Ppd-D1* promoter as detailed in Table 2.1.

The quality of the RNA-seq reads was assessed using FASTQC (Andrews, 2018). The mean quality scores at all read positions had Phred scores greater than 28 and were, therefore, determined to be 'good', with a small decrease in quality at the end of the read (Figure 4.1). This degradation of quality scores is expected and is often the result of phasing (or pre-phasing) during sequencing by synthesis, where reads become 'out-of-step' with one another (ecSeq Bioinformatics, 2017). This is frequently due to errors in removing terminator caps during sequencing cycles, meaning a single nucleotide may be read twice, or a nucleotide is skipped. This issue is widely described (ecSeq Bioinformatics, 2017), and so was deemed to be acceptable.

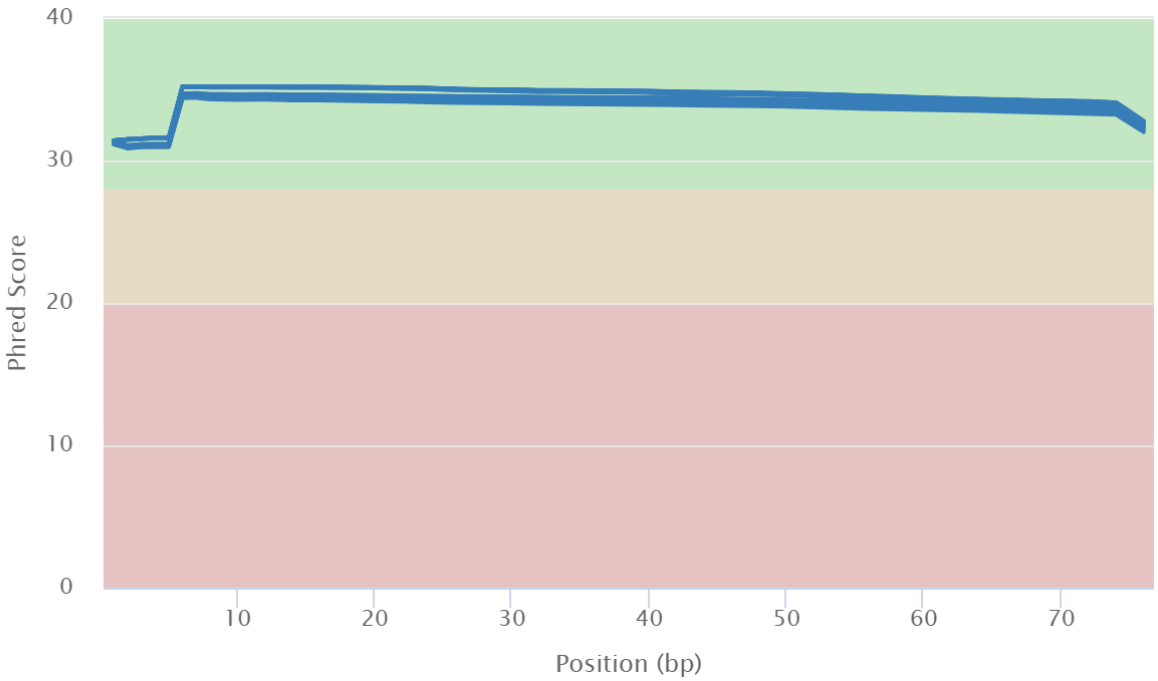

Figure 4.1: The mean quality scores of all raw RNA-seq reads at each read position. Each line represents the forward or reverse reads for a single sample. Green shading represents 'good' Phred scores (>28), orange shading represents 'reasonable' scores (20-28) and red represents 'poor' scores (<20). Figure produced using FASTQC (Andrews, 2018).

The overall quality score of each sequence was also analysed (Figure 4.2). Most of the reads had a 'good' mean quality score. This analysis, along with the data presented in Figure 4.1, provides evidence that all the samples had a similar, good quality profile.

The GC content of the reads was analysed using FastQC and all samples failed the FastQC GC content module (Figure 4.3). The samples showed a positively skewed distribution, as opposed to the expected normal distribution. Also, several small peaks were observed, including one only seen in reverse reads (indicated in Figure 4.3 by an arrow). Additional peaks are often an indication of contamination, such as adapter dimers. The peak at 48% matches the estimated GC content for wheat cDNA assemblies of 47.74% (Brenchley *et al*., 2012). Although the samples failed this FastQC module, this was determined to be acceptable. The sequencing libraries passed all quality control checks (data not shown), so there was unlikely to be any contamination at the library preparation stage. However, there is the possibility of that contamination occurred subsequently during the sequencing workflow. Also, RNA-seq studies can produce skewed GC-content distributions due to the presence of naturally occurring overrepresented sequences. FastQC assumes the sequencing libraries are random, which RNA-seq libraries will not be.

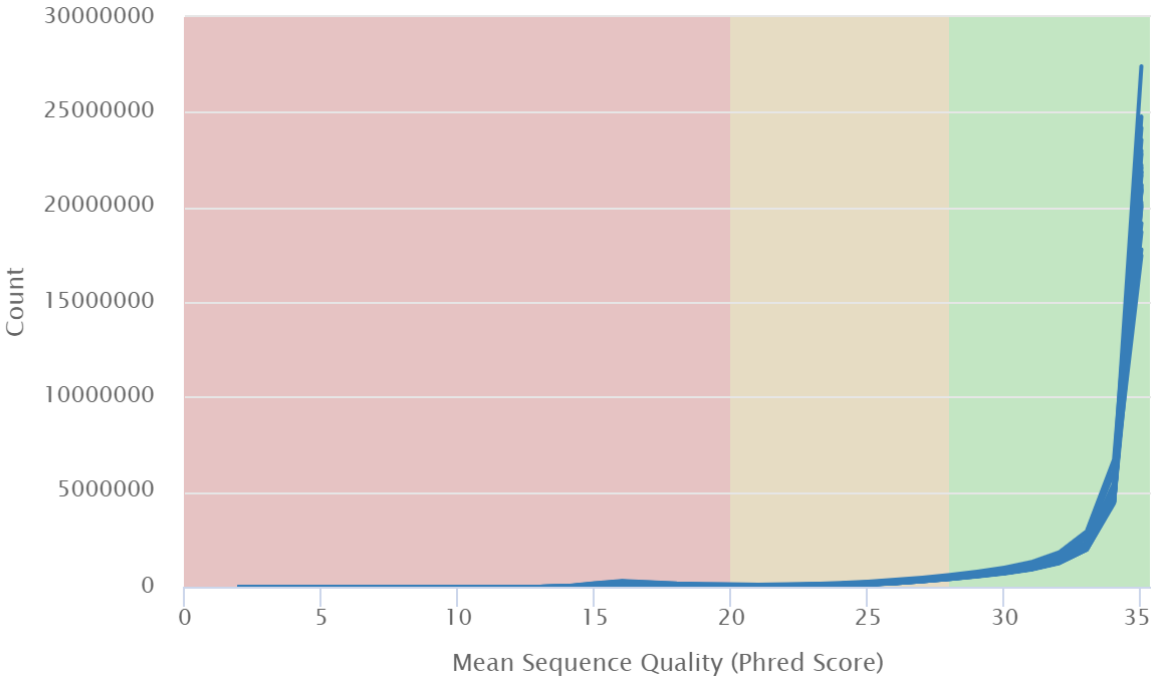

Figure 4.2: The number of raw RNA-seq reads with each mean sequence quality score. Each line represents the forward or reverse reads for a single sample. Green shading represents 'good' Phred scores (>28), orange shading represents 'reasonable' scores (20-28) and red represents 'poor' scores (<20). Figure produced using FASTQC (Andrews, 2018).

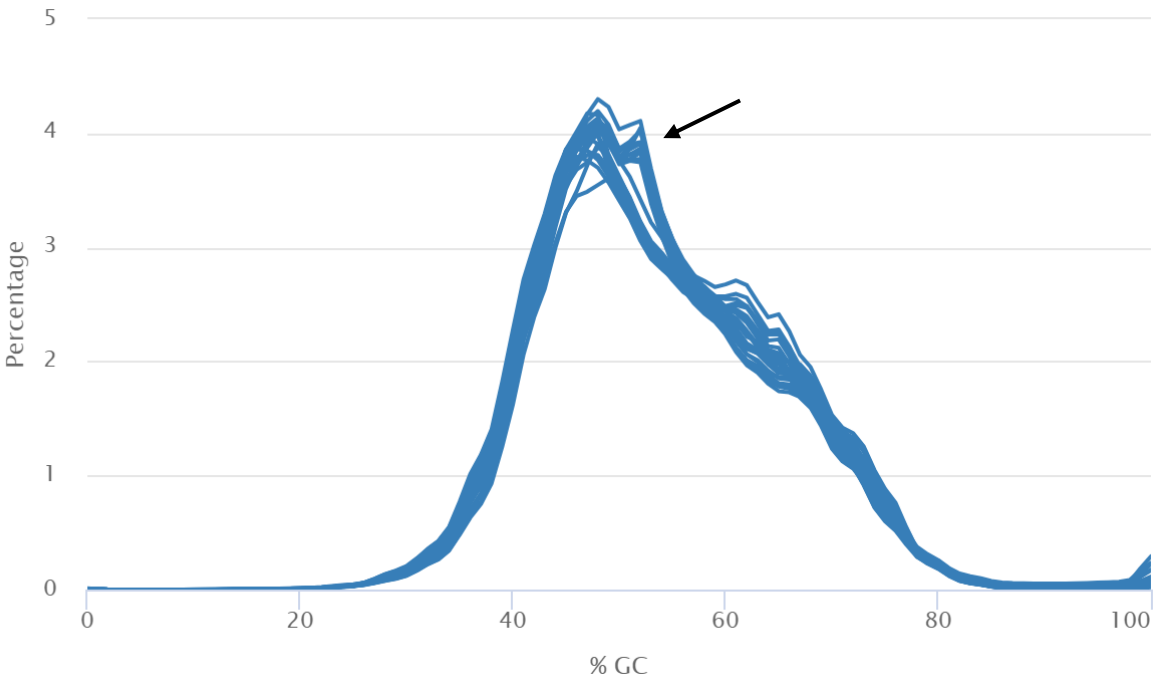

Figure 4.3: The percentage of reads per sample with each average GC content value. Lines represent either the forward or reverse reads for a single sample. The arrow indicates a peak which is present in the graphs generated for reverse reads and absent in forward reads. Figure produced using FASTQC (Andrews, 2018).

Attempts to improve the quality of the reads were made by trimming them according to a database of Illumina adapters (provided by Dr Alex Paterson (University of Bristol Genomics Facility; Bristol, UK)). The trimmed reads were again assessed using FastQC, and no substantial improvements were made on the quality of the reads for any FastQC module. Therefore, the raw reads were used to avoid inadvertently losing any potentially informative data.

4.2 RNA-SEQ RESULTS

4.2.1 *ubi1* Experiment

Based on the results of the DeSeq2 analysis (as described in section 2.6.4) (Love *et al*., 2014), there was a statistically significant increase in the expression of both SunTag constructs VP64-ScFv (1.89-fold increase, $p < 0.001$, Figure 4.4) and dCas9-SunTag$_{x10}$ (1.65-fold increase, $p = 0.012$, Figure 4.5) in B3598.6 protoplasts transformed with *ubi1* sgRNAs compared to non-transformed B3598.6 protoplasts. There was no statistically significant change in the expression of dCas9-SunTag$_{x10}$ or VP64-ScFv in B3598.6 protoplasts when transformed with *Ppd-D1* sgRNAs. There was a statistically significant difference in the expression of VP64-ScFv in B3598.6 protoplasts transformed with *ubi1* sgRNAs compared to *Ppd-D1* sgRNAs (1.65-fold change, $p = 0.017$). However, there was no statistically significant difference in the expression of dCas9-SunTag$_{x10}$ in B3598.6 protoplasts transformed with

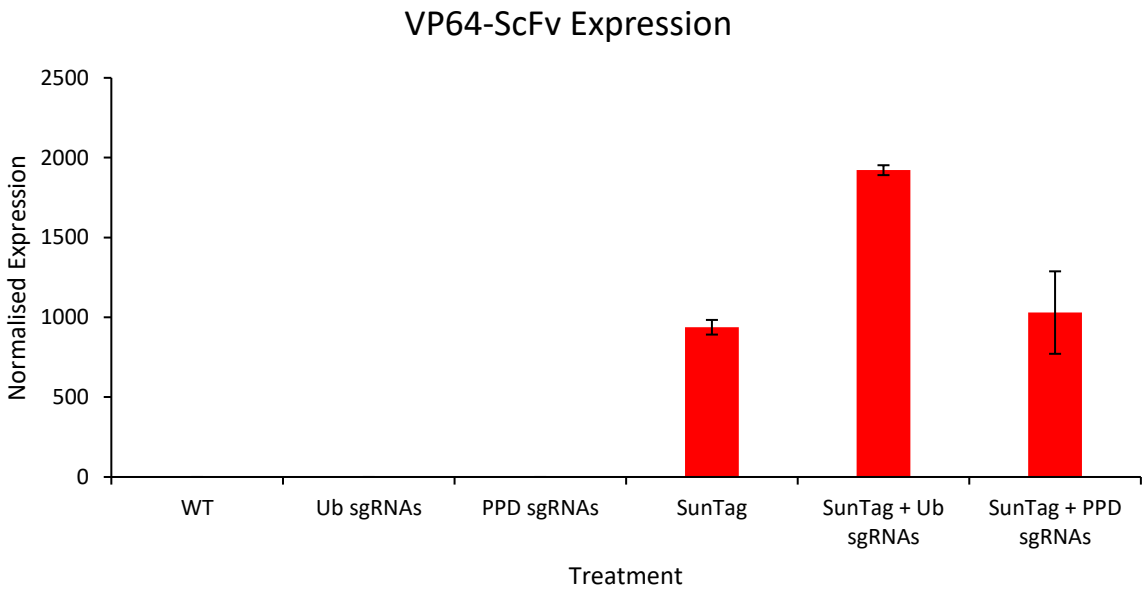

Figure 4.4: Normalised expression levels of VP64-ScFv in protoplasts generated from WT wheat and line B3598.6 which contains the SunTag constructs. Normalised expression level data was generated by DeSeq2 as described in section 2.6.4 (Love *et al*., 2014). Protoplasts were transformed with sgRNAs targeting the *ubi1* promoter or the *Ppd-D1* promoter. Error bars represent the standard error of the mean.

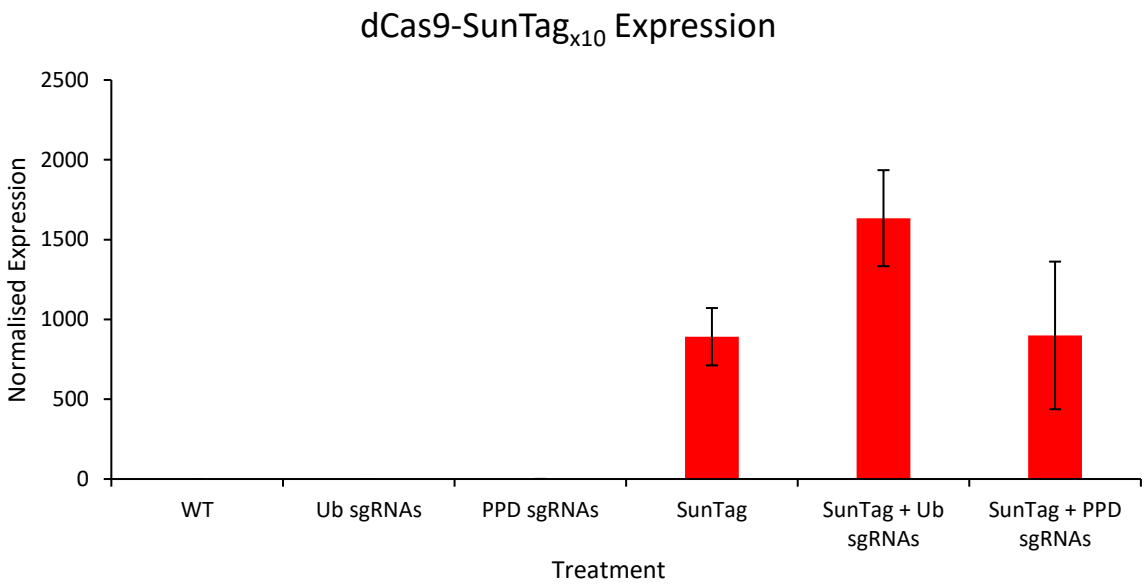

Figure 4.5: Normalised expression levels of dCas9-SunTag$_{x10}$ in protoplasts generated from WT wheat and line B3598.6 which contains the SunTag constructs. Normalised expression level data was generated by DeSeq2 as described in section 2.6.4 (Love *et al*., 2014). Protoplasts were transformed with sgRNAs targeting the *ubi1* promoter or the *Ppd-D1* promoter. Error bars represent the standard error of the mean.

ubi1 sgRNAs compared to *Ppd-D1* sgRNAs ($p = 0.874$). The expression data for VP64-ScFv suggests that the CRISPRa system can be used to increase gene expression and is specific according to the sgRNAs used. The expression data for dCas9-SunTag$_{x10}$ are inconclusive.

4.2.2 *Ppd-D1* Experiment

No statistically significant difference in *Ppd-D1* expression was detected in cells transformed with *Ppd-D1* sgRNAs. The normalised expression level of *Ppd-D1* in cells containing the SunTag constructs and sgRNAs targeting *Ppd-D1* indicated that there were no *Ppd-D1* transcripts present (Figure 4.6). The adjusted *p*-value for *Ppd-D1* was excluded from the results of the DeSeq2 analysis as a result of independent filtering. Independent filtering aims to increase the power of the statistical analysis by ensuring all adjusted *p*-values satisfy a false discovery rate threshold. This is achieved by excluding genes with very low levels of expression. The expression level of *Ppd-D1* was determined by DeSeq2 (Love *et al*., 2014) to be too low for conclusions to be drawn on any changes in expression.

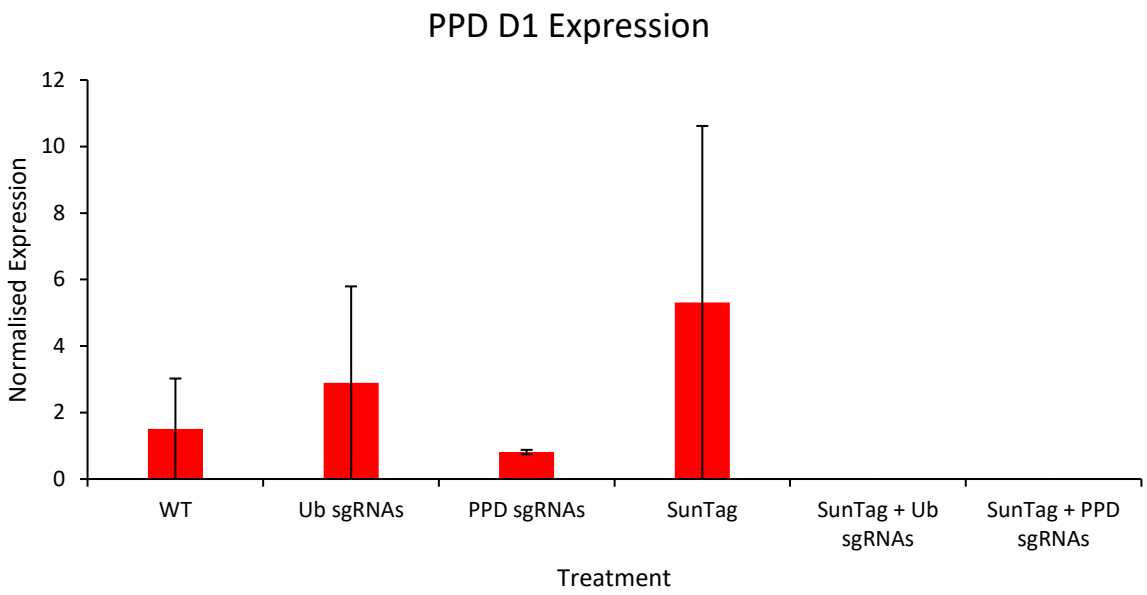

Figure 4.6: Normalised expression levels of *Ppd-D1* in protoplasts generated from WT wheat and line B3598.6 which contains the SunTag constructs. Normalised expression level data was generated by DeSeq2 as described in section 2.6.4 (Love *et al*., 2014). Protoplasts were transformed with sgRNAs targeting the *ubi1* promoter or the *Ppd-D1* promoter. Error bars represent the standard error of the mean.

4.2.3 SunTag System Specificity

It is important to estimate any off-target effects of the SunTag system if it were to become a useful tool, as it would need to be both powerful and specific. The principal component analyses in Figures 4.7 and 4.8 show the global effect on the transcriptome of transforming cells with the two SunTag constructs and sgRNAs. In both studies the groups expressing the SunTag constructs cluster together. 2733 genes were differentially expressed in cells containing SunTag compared to WT cells. 1158 (42%) of these were downregulated, and 1575 (58%) were upregulated compared to the WT.

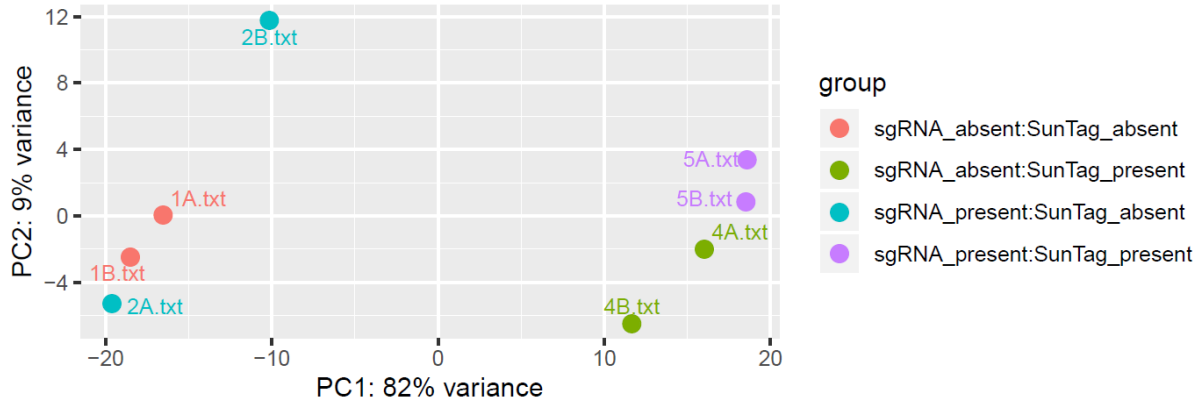

Figure 4.7: Principle component analysis (PCA) of RNA-seq data from WT and B3598.6 protoplasts transformed with sgRNA constructs targeting the *ubi1* promoter. The 'group' indicates the treatments applied to each sample. There were two replicates per treatment. The percentages assigned to each axis refer to the percentage of the data described by that principle component. Figure produced using DeSeq2 (Love *et al*., 2014).

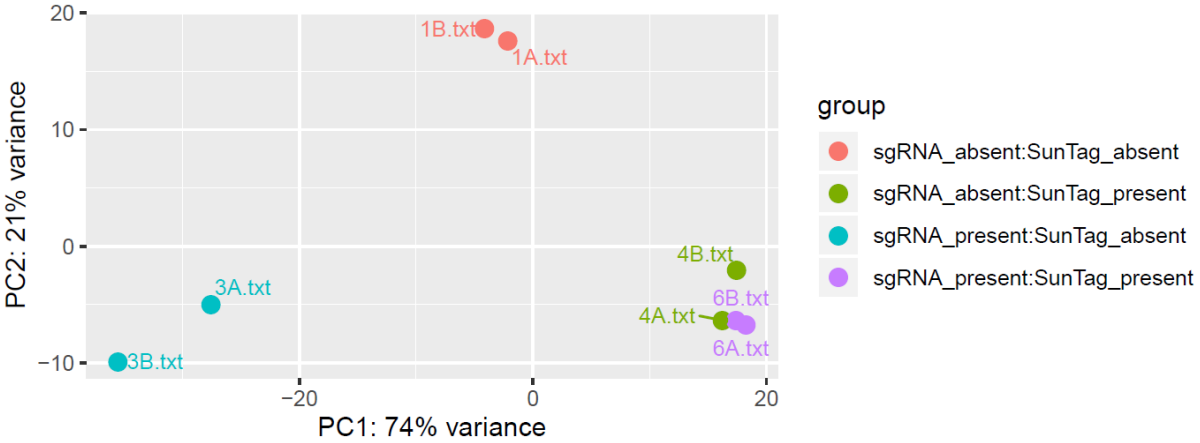

Figure 4.8: Principle component analysis (PCA) of RNA-seq data from WT and B3598.6 protoplasts transformed with sgRNA constructs targeting the *Ppd-D1* promoter. The 'group' indicates the treatments applied to each sample. There were two replicates per treatment. The percentages assigned to each axis refer to the percentage of the data described by that principle component. Figure produced using DeSeq2 (Love *et al*., 2014).

15,042 genes were differentially expressed in B3598.6 cells containing the SunTag constructs and *ubi1* sgRNAs compared to WT cells. A BLAST (Basic Local Alignment Search Tool) search conducted by Dr Gary Barker (University of Bristol; Bristol, UK) showed that only 7 of these genes shared sequence similarity (maximum of one mismatch) with the *ubi1* sgRNAs, none of which were within the optimum range for CRISPRa sgRNA design. Compared to WT cells 3,881 genes were differentially expressed in

cells containing the SunTag constructs and *Ppd-D1* sgRNAs. Of these differentially expressed genes, 22 shared some sequence similarity (maximum of one mismatch) with the *Ppd-D1* sgRNAs. Only one of these matching motifs was in the optimum range for the CRISPRa system. The differentially expressed genes were located on all chromosomes in all genomes and had a range of basal expression levels.

4.3 INTERPRETATION OF THE RESULTS

This study was unable to replicate the results of at least a 10-fold increase in gene expression reported by Tanenbaum *et al.* (2014) in human cells. This may suggest that the SunTag system is not as powerful when used in wheat, however, the lack of any observed gene upregulation may be the result of mistargeted sgRNAs. Gilbert *et al.* (2014) suggested that designing sgRNAs to target -400 to -50 bp from the transcription start site induces the greatest average level of upregulation for the SunTag system. The *ubi1* sgRNAs were designed to target the optimal range at positions -121 bp, -148 bp, -235 bp, and -258 bp (Figure 2.2, Table A.2), while the *Ppd-D1* sgRNAs were targeted to sequences between the transcription start site and the translation start site ranging from positions +68 bp to +552 bp (Figure 2.2, Table A.2). Therefore, the TADs of the SunTag system were not targeted to the optimum range. This could be an alternative reason that transformation of B3598.6 with *Ppd-D1* sgRNAs resulted in no change in *Ppd-D1* gene expression. The SunTag complex may have physically interfered with transcription, inhibiting mRNA production. Further work could be undertaken with sgRNAs designed according to the suggestions made by Gilbert *et al*. (2014), or ideally, a separate preliminary study should be undertaken to determine the optimal sgRNA target range in wheat. Also, Ramírez-González *et al.* (2018) reported that the base level of expression of *Ppd-D1* is very low at just 4.72 TPM. This means that any small increase in gene expression would have been difficult to detect. In future work, a higher number of replicates or a more sensitive technique could be used, for instance, qRT-PCR (quantitative RT-PCR). However, this technique could not be used to assess off-target effects, as it is used to quantify the expression of a single gene.

The results of the PCA suggest that the insertion of SunTag constructs affects gene expression globally, even in the absence of sgRNAs. Almost none of the differentially expressed genes showed any sequence similarity to the sgRNAs used, so the change in expression was unlikely to have been due to imperfect binding between the sgRNAs and endogenous DNA. This high number of differentially expressed genes was not observed by other studies using the SunTag CRISPRa system (Konermann *et al*., 2015), or studies using dCas9 to repress genes (Gilbert *et al*., 2013a). This is a surprising result which should be investigated further, as all the treated cells showed unexpected differential expression patterns compared to the WT. However, some of the differentially expressed genes may have been a result of the lack of replicates in this study (two per treatment). This means

that any stochasticity in gene expression would have had a large impact on the analysis. Therefore, further studies with more replicates should be carried out.

4.4 LIMITATIONS OF THE RNA-SEQ ANALYSIS

The number of replicates in this study was two per treatment, however, Schurch *et al.* (2016) suggested that three replicates are the minimum required for RNA-seq and that 12 is ideal. This was supported by the high standard errors of the mean for the data points in Figures 4.4, 4.5 and 4.6; these reflect the natural variation in gene expression which may have obscured patterns caused by the treatments applied. Therefore, the number of replicates in this study was not sufficient to determine conclusively whether there was upregulation of *Ppd-D1* expression in B3598.6 protoplasts treated with *Ppd-D1* sgRNAs. Further RNAseq studies should include a greater number of replicates to assess the viability of this CRISPRa system.

The tissue used to generate the protoplasts came from heterogeneous seed stock due to the unavailability of a homozygous line. Therefore, not all cells within the B3598.6 protoplast pool contained the SunTag constructs, so the CRISPRa system would have only worked in a proportion of the cells in the sample (43.75% according to the genotyping described in section 3.1.3). A non-segregating seed stock should be generated to fairly assess the effectiveness of this system in wheat. Additionally, not all cells would have been transformed with the sgRNAs. Therefore, any small increase in gene expression in transformed cells may be undetectable. Furthermore, any detected increases in expression are likely to be underestimates of the potential of the system. A temporary solution to this issue would be fluorescence activated cell sorting (FACS) of the protoplasts, as transformant cells are tagged with sfGFP and YFP.

Future work would allow the SunTag system to be assessed more comprehensively. It has been noted in previous studies that there can be a large amount of variation in the level of upregulation of gene expression achieved, even when targeting the same gene (Maeder *et al*., 2013). It has been proposed that this is the result of competitive binding of endogenous transcription factors, variation in the level of binding the SunTag complex achieves with the DNA (Maeder *et al*., 2013), the endogenous epigenetic state of the target DNA, or the base level of gene expression (Chavez *et al*., 2015). These effects could be controlled in multiple ways. For instance, targeting a single endogenous gene is not ideal, as there may be a specific reason why the system would be less effective when targeted to it. For instance, *Ppd-D1* is partially regulated by the circadian clock (Kiseleva *et al*., 2017; Boden *et al*., 2015), meaning the expression level would be affected by the time at which the RNA was extracted. Ideally, a range of genes should be targeted with different endogenous expression patterns and in a range of regulatory networks. Genes with few or no known downstream targets would be

excellent candidates for assessing the off-target effects of the system, as increases in the expression of downstream genes can be confused with non-specific upregulation. Additionally, genes with low, medium, and high base levels of expression should be targeted to assess the effect of endogenous gene expression on the level of upregulation. It must also be considered that some genes may be components of regulatory loops, meaning that their expression is limited by other genes. Therefore, the level of possible upregulation of a target gene may be limited without altering the expression of other genes in the network.

Given these factors which can affect the level of upregulation possible, it is also important to account for differences in the effectiveness of the system in different tissues of wheat plants, or at different developmental stages. The promoters driving the expression of SunTag constructs may not act constitutively in all tissues at all times (Chapter 3.1.2), and the same may be the case for the promoter driving the sgRNA constructs. Therefore, ideally, RNA should be sampled from various tissues at various stages of development to assess the effectiveness of the CRISPRa system in the entire plant, not only in protoplasts generated from young leaf tissue.

A further extension of this work would be to assess the phenotypic effect induced upregulation has on the plants. This has been used to assess the effectiveness of SunTag in *A. thaliana* where the system was used for targeted DNA methylation (Papikian *et al*., 2019). This would most likely require plants stably transformed with both the SunTag constructs and sgRNAs to measure any change in phenotype.

4.5 CONCLUSIONS

This study showed that the SunTag CRISPRa system may increase expression of exogenous genes which form part of a feedback loop in wheat protoplasts by a mean of 1.77-fold. However, no increase in gene expression was observed when the SunTag system was targeted to endogenous gene *Ppd-D1*.

These results suggest that although the SunTag system is functional, it is not capable of high levels of gene upregulation in wheat protoplasts. Further work would determine whether the system can be usefully applied in wheat.

CHAPTER 5: DISCUSSION

5.1 SUMMARY OF FINDINGS

5.1.1 Aim 1: Implement transient and stable expression of a modified SunTag system in wheat cells and plants

Transformation of protoplasts and plants with the two SunTag constructs, dCas9-SunTag$_{x10}$ and VP64-ScFv, was achieved. Transient expression in protoplasts was confirmed using fluorescent microscopy to observe GFP tags. Stable transformation *in planta* was confirmed using PCR. The transformation rate was lower than predicted for both transient and stable transformation. The hypotheses to explain the low transformation rate include weak construct expression, and the constructs having deleterious effects in wheat.

5.1.2 Aim 2: Verify that the dCas9-SunTag$_{x10}$ and VP64-ScFv SunTag constructs are present in wheat cells and plants, and are expressing the corresponding mRNA and proteins. Verify that these SunTag proteins are interacting

The presence of correctly processed mRNA was verified in protoplasts and stably transformed plants using RT-PCR. The presence of the VP64-ScFv protein was shown through LC-MS analysis of co-IP products. The presence of the dCas9-SunTag$_{x10}$ protein was implied based upon the co-IP LC-MS results; the presence of the VP64-ScFv protein was reliant on the presence of dCas9-SunTag$_{x10}$ protein. The SunTag proteins were also strongly suggested to have been interacting based on the co-IP LC-MS study.

5.1.3 Aim 3: Induce specific gene expression of both inserted and endogenous genes, and quantify any changes in gene expression

The system was shown to significantly upregulate the *ubi1* promoter with a mean 1.77-fold increase in SunTag construct expression. When sgRNAs targeting *Ppd-D1* were used, no increase in *Ppd-D1* expression was detected.

The *ubi1* study involved a positive feedback loop, which was likely to increase the level of upregulation achieved. Therefore, one hypothesis to explain this result is that the level of upregulation possible without this positive feedback loop is too low to be detected. The base level of *Ppd-D1* expression was low, making any increase in expression more difficult to identify. Further optimisation of the CRISPRa system is necessary to induce increased gene expression in endogenous genes.

5.2 LIMITATIONS OF THE STUDY

Although all the objectives of this study were fulfilled, there were some issues which require further investigation.

The evidence of the presence of the dCas9-SunTag$_{x10}$ protein, as well as its interaction with the VP64-ScFv protein, was based on assumptions made about the co-IP LC-MS analysis. This falls short of conclusive evidence, however, the RNA-seq data strongly suggests that both these proteins are present and interacting as the CRISPRa system should not work unless this is the case. Further work could be completed to provide stronger evidence; for instance, by optimising the co-IP reaction to reduce the number of non-specific proteins in the final sample; this would improve the likelihood of detecting the dCas9-SunTag$_{x10}$ protein.

One of the main limitations of the RNA-seq study was the lack of replicates; there were only two per treatment. There was considerable stochasticity between replicates. This reduces the power of any statistical analysis carried out on the data and so smaller changes in expression may not be detected.

The other main limitation was that not all cells within a sample had the same genotype. At the time of the study setup, no non-segregating line transformed with the SunTag constructs had been generated. Therefore, the shoots from which the protoplasts were isolated would have been a mix of positive and negative genotypes. Also, the sgRNA plasmids would not have transformed 100% of cells. This makes it difficult to estimate the level of upregulation possible using the CRISPRa system and means that more subtle effects may have been undetectable.

5.3 FURTHER WORK

5.3.1 Optimising the System

It is important to note that the system was only implemented *in vitro*, and so the next logical step would be to apply it *in planta*. The SunTag constructs have been shown to express in stably transformed plants (section 3.2.3). sgRNAs would also need to be either transiently or stably expressed in these plants. If stable expression were required, the sgRNA constructs would need to be inserted into the genome using *Agrobacterium tumefaciens*-mediated or biolistic techniques. However transient expression could also be induced very simply with a novel transformation technique using nanoparticles, as described by Doyle *et al*. (2019). Implementing the system *in planta* would allow a more accurate estimation of the power of this CRISPRa system, as all cells should contain the SunTag constructs. All cells would contain sgRNAs if plants were stably transformed with the sgRNA constructs. This may mean that any small increase in gene expression induced by the SunTag system would be detectable.

There is enormous potential for development beyond the studies described here. Initially, the system needs to be optimised for use in wheat, as at present it is not capable of upregulating at least some endogenous genes. The length of the amino acid linkers between GCN_4 epitopes has been shown to strongly affect the activity of the associated effector molecules (Morita *et al*., 2016). For instance, in a study by Morita *et al*. (2016) where gene activation was achieved using the DNA demethylator TET1 (ten-eleven translocation methylcytosine dioxygenase 1), the standard SunTag system showed no improvement over a simple protein fusion. However, by optimising linker length, demethylation rates were increased from 14% to 38% (Morita *et al*., 2016). This optimisation strategy could be implemented to improve the CRISPRa system discussed here. Other variables should also be investigated and optimised, such as the number of GCN_4 epitopes in the SunTag array. TADs other than VP64 could be used, such as VPR (Chavez *et al*., 2015) or TV (Li *et al*., 2017). A recent study tested different CRISPRa techniques in *N. benthamiana*; SunTag was found to not be the most effective CRISPRa system (Selma *et al*., 2019). scRNA-gRNA2.1 combined with dCas9:EDLL-MS2:VPR induced the greatest level of upregulation (Selma *et al*., 2019). A similar optimisation study could be conducted in wheat to ascertain which approach is most effective, as variation in transcriptional machinery and promoter structure in different species may affect which system is most powerful and specific. This work would help to increase the power of the system sufficiently to be capable of upregulating endogenous genes.

It is also important to consider whether spCas9 is the most efficient Cas9 orthologue to use. As mentioned in section 1.1, comparative studies have shown that there may be more effective or specific variants than spCas9, even though it is the most used nuclease (Gilbert *et al*., 2013a). Orthologues of dCas9 from species other than *S. pyogenes* have been used for fluorescent labelling (Chen *et al*., 2016). In addition, Rock *et al.* (2017) screened multiple dCas9 orthologues for use in *Mycobacterium tuberculosis* to decrease gene expression. They found that the dCas9 orthologue from *Streptococcus thermophilus* (CRISPR1 variant) was over 20 times more efficient than the *S. pyogenes* orthologue (Rock *et al*., 2017). This shows the importance of choosing the correct orthologue for a target species. sgRNA design is also vital, as shown by Gilbert *et al*. (2014), who found that the optimal position for CRISPRa sgRNAs is -400 to -50 bp upstream of the transcription start site. A similar study could be conducted in wheat to ensure this optimal window is the same as Gilbert *et al.* (2014) observed in human cells, as the different transcriptional machinery in plants may alter this.

Another factor to be investigated is the number of locations the system is used to target. SunTag has been successfully used to target multiple genes concurrently by multiplexing sgRNAs (Papikian *et al*., 2019); this could be tested in wheat. The polyploid nature of wheat could be interrogated using this technique; it would be possible to target one homeologue or all three. For instance, in the case of the

Ppd-D1 study described here, only the D-genome copy was targeted. In future, the effect of targeting additional gene copies on the A and B genome could be assessed. In some cases, certain copies of a gene are more influential on the phenotype than others, in which case targeting different combinations of homeologues would influence the level of upregulation observed. By using CRISPRa to upregulate each homeologue separately, the influence of each gene copy could also be investigated.

5.3.2 Creating a More Flexible System

There is scope for making the SunTag CRISPRa system more flexible, meaning that it could be applied to a greater number of research questions. The constructs could be driven by an inducible promoter instead of a constitutively expressing one. This includes light-inducible systems, which have already been used to drive the SunTag system in human cells (Polstein & Gersbach, 2015). Heat-shock promoters such as the barley (*Hordeum vulgare*) HVHSP17 promoter (Freeman *et al*., 2011) could also be used. Tissue-specific promoters could be used, such as the sedoheptulose-1,7-bisphosphatase (SBPase) promoter from *Brachypodium distachyon*, which has been shown to express genes specifically in wheat leaf tissue (Alotaibi *et al*., 2018). These strategies may reduce the already very low level of SunTag construct expression and limit the upregulation possible even further. On the other hand, by only expressing the SunTag constructs when necessary, any deleterious effects of the constructs may be limited.

In the long-term, to create the most flexible system possible, 'CRISPRa-ready' crops could be created. This idea has been proposed by Hickey *et al*. (2019) for Cas9-editing, however, the same theory could be applied to dCas9 expression manipulation. This would involve generating plants which stably express the dCas9-SunTag$_{x10}$ and VP64-ScFv constructs. These 'CRISPRa-ready' lines could then be transformed with sgRNAs and upregulation of any target could be quickly achieved. If more flexibility was necessary, the lines could only include the dCas9-SunTag$_{x10}$ construct, and an ScFv with any effector domain could be used to transform the plant; for instance, gene expression downregulators could be used, or epigenetic modifiers such as TET1.

5.4 APPLICATIONS OF CRISPRa IN WHEAT

Applying this CRISPRa technique in commercial crops would encounter issues regarding regulation in the European Union (EU). In 2018 the Court of Justice of the EU ruled that gene editing is subject to the same restrictions that apply to genetically modified organisms (GMOs) (Court of Justice of the European Union, 2018), which are very restrictive; no GMO varieties are commercially grown in the United Kingdom (United Kingdom Government, 2019). This policy is likely to be in place in the short term, although may change at the end of the transition period

(1st January 2021) when the UK may stop following EU legislation on GMOs. In the United States, although Cas9-edited crops are not regulated if they "could otherwise have been developed through traditional breeding techniques" (United States Department of Agriculture Animal and Plant Health Inspection Service, 2012), the CRISPRa system described in this study would require regulatory approval as the plants are stably transformed with foreign genes. Therefore, under current regulatory conditions, the commercial potential of this CRISPRa technique in at least two major markets (the EU and the United States) is limited. However, there is the potential of using GM-free techniques such as ribonucleoproteins (RNPs) (complexes of RNA and RNA-binding proteins) for short term manipulation of gene expression. This technique would involve inserting complete complexes consisting of a dCas9-SunTag$_{x10}$ protein, ten VP64-ScFv proteins, and an sgRNA. These complexes would then directly upregulate target genes. This transient technique could theoretically be used without requiring stringent regulation, with the advantage of a lower level of off-target activity (Liang *et al*., 2017). RNPs have been successfully used in wheat for basic CRISPR-Cas9 editing (Liang *et al*., 2017). A possible application of this method of short-term upregulation of specific genes would be during droughts to increase crop drought tolerance, meaning that any disadvantages imposed by the increased tolerance, such as yield decreases, are minimised.

The greatest current potential of the CRISPRa system is as a research tool. As stated in section 1.2, CRISPRa can be used in gain-of-function studies as well as overexpression screens (Konermann *et al*., 2015), which would be extremely valuable in fundamental wheat research. Studies using CRISPRa could be used to identify candidate alleles for conventional breeding, allowing the technology to directly benefit commercial crop development.

5.4 CONCLUSIONS

The development of a new tool for wheat research and breeding provides a boost to the global effort to increase yields to match increasing global demand. This study's aim was to translate a technique developed in model organisms into a crop species. No CRISPRa system has so far been implemented in wheat. The SunTag system is a second-generation technique which has been shown to be more effective than dCas9-VP64 in previous studies (Tanenbaum *et al*., 2014).

In this study, it has been shown that the components of the SunTag CRISPRa system can be expressed in wheat. It has also been shown to be capable of upregulation through a positive-feedback pathway. However, for the endogenous gene the system was tested on, no increase in expression was detected. Although this may be due to the specific gene targeted, this result suggests that a significant amount of further work would be needed to optimise this system for use in wheat for it to become a useful technique. There is scope for improvement, including using different TADs, as well as recruiting more

TAD copies using different techniques (such as sgRNA 2.0 (Konermann *et al*., 2015)). There is also the potential of improving the flexibility of CRISPRa by using specific or inducible promoters, or through the development of a 'CRISPRa-ready' wheat line.

This project achieved almost all the aims as set out in section 1.7. Unfortunately, the results of this study mean that it falls short of proving the concept works in wheat. However, it does show that there is potential for this system to function in this important crop species and become a useful tool for researchers and breeders alike. Improving the tools used for crop development is a vital part of the solution to global food insecurity. Increasing wheat yields is of particular significance, as it is the most widely grown crop globally and is a staple crop particularly in the developing world (Shiferaw *et al*., 2013). With further development, CRISPRa could provide another tool with which to tackle global food insecurity in the 21st century.

BIBLIOGRAPHY

Abcam. (2010) *Protocols book*. Available from: https://docs.abcam.com/pdf/misc/abcam-protocols-book-2010.pdf [Accessed 2nd March 2019].

Alotaibi, S. S., Sparks, C. A., Parry, M. A. J., Simkin, A. J. & Raines, C. A. (2018) Identification of leaf promoters for use in transgenic wheat. *Plants.* 7 (2), 27. Available from: doi:10.3390/plants7020027.

Andrews, S. (2018) *FastQC*. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed 20th August 2019].

Behura, S. K. & Severson, D. W. (2013) Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. *Biological Reviews.* 88 (1), 49-61. Available from: doi:10.1111/j.1469-185X.2012.00242.x.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J. & Gluenz, E. (2017) A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *Royal Society Open Science*, 4, 170095. Available from: doi:10.1098/rsos.170095.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F. & Marraffini, L. A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research.* 41 (15), 7429-7437. Available from: doi:10.1093/nar/gkt520.

Biłas, R., Szafran, K., Hnatuszko-Konka, K. & Kononowicz, A. K. (2016) *Cis*-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue and Organ Culture.* 127, 269-287. Available from: doi:10.1007/s11240-016-1057-7.

Boden, S. A., Cavanagh, C., Cullis, B. R., Ramm, K., Greenwood, J., Finnegan, E. J., Trevaskis, B. & Swain, S. M. (2015) *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants.* 1, 14016. Available from: doi:10.1038/nplants.2014.16.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* 72 (1-2), 248-254. Available from: doi:10.1016/0003-2697(76)90527-3.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W. R., Hall, A., Mayer, K. F. X., Edwards, K. J., Bevan, M. W. & Hall, N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature.* 491, 705-710. Available from: doi:10.1038/nature11650.

Carroll, D., Morton, J. J., Beumer, K. J. & Segal, D. J. (2006) Design, construction and *in vitro* testing of zinc finger nucleases. *Nature Protocols,* 1 (3), 1329-1341. Available from: doi:10.1038/nprot.2006.231.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J. & Voytas, D. F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research.* 39 (12), e82. Available from: doi:10.1093/nar/gkr218.

Chan, V., Dreolini, L. F., Flintoff, K. A., Lloyd, S. J. & Mattenley, A. A. (2002) The effect of increasing plasmid size on transformation efficiency in *Escherichia coli. Journal of Experimental Microbiology and Immunology.* 2, 207-223. Available from: https://microbiology.ubc.ca/sites/default/files/roles/drupal_ungrad/JEMI/2/2-207.pdf [Accessed 3rd April 2020].

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., Iyer, E. P. R., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J. & Church, G. M. (2015) Highly efficient Cas9-mediated transcriptional programming. *Nature Methods.* 12 (4), 326-328. Available from: doi:10.1038/nmeth.3312.

Chavez, A., Tuttle, M., Pruitt, B. W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S. J., Cecchi, R. J., Kowal, E. J. K., Buchthal, J., Housden, B. E., Perrimon, N., Collins, J. J. & Church, G. (2016) Comparison of Cas9 activators in multiple species. *Nature Methods.* 13 (7), 563-567. Available from: doi:10.1038/nmeth.3871.

Chen, B., Hu, J., Almeida, R., Liu, H., Balakrishnan, S., Covill-Cooke, C., Lim, W. A. & Huang, B. (2016) Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Research.* 44 (8), e75. Available from: doi:10.1093/nar/gkv1533.

Cheng, A. W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A. & Wang, H. (2016) Casilio: A versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Research.* 26, 254-257. Available from: doi:10.1038/cr.2016.3.

Christensen, A. H., Sharrock, R. A. & Quail, P. H. (1992) Maize polyubiquitin genes: Structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Molecular Biology.* 18, 675-689. Available from: doi:10.1007/BF00020010.

Clarivate Analytics. (2020a) *Web of Science*. Available from: http://apps.webofknowledge.com [Accessed 18th February 2020].

Clarivate Analytics. (2020b) *Web of Science*. Available from: http://apps.webofknowledge.com [Accessed 26th March 2020].

Comai, L. (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics.* 6, 836-846. Available from: doi:10.1038/nrg1711.

Court of Justice of the European Union. (2018) *Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive*. Available from: https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-07/cp180111en.pdf [Accessed 22nd November 2019].

Cribbs, A. P. & Perera, S. M. W. (2017) Science and bioethics of CRISPR-Cas9 gene editing: An analysis towards separating facts and fiction. *The Yale Journal of Biology and Medicine.* 90 (4), 625-634. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5733851/#__ffn_sectitle [Accessed 29th March 2020].

Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. (1961) General nature of the genetic code for proteins. *Nature.* 192 (4809), 1227-1232. Available from: doi:10.1038/1921227a0.

Dai, S., Zheng, P., Marmey, P., Zhang, S., Tian, W., Chen, S., Beachy, R. N. & Fauquet, C. (2001) Comparative analysis of transgenic rice plants obtained by *Agrobacterium*-mediated transformation and particle bombardment. *Molecular Breeding: New Strategies for Plant Improvement.* 7 (1), 25-33. Available from: doi:10.1023/a:1009687511633.

DeFrancesco, L. (2011) Move over ZFNs. *Nature Biotechnology.* 29 (8), 681-684. Available from: doi:10.1038/nbt.1935.

Designing Future Wheat. (2020) *Designing Future Wheat.* Available from: https://designingfuturewheat.org.uk/ [Accessed 1st April 2020].

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 29 (1), 15-21. Available from: doi:10.1093/bioinformatics/bts635.

Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D. & Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal 'skip'. *Journal of General Virology.* 82 (5), 1013-1025. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-82-5-1013#abstract_content [Accessed 4th April 2020].

Doyle, C., Higginbottom, K., Swift, T. A., Winfield, M., Bellas, C., Benito-Alifonso, D., Fletcher, T., Galan, M. C., Edwards, K. & Whitney, H. M. (2019) A simple method for spray-on gene editing *in planta*. *bioRxiv.* [Preprint] Available from: doi:10.1101/805036.

Dvořák, J., di Terlizzi, P., Zhang, H. & Resta, P. (1993) The evolution of polyploid wheats: Identification of the A genome donor species. *Genome.* 36 (1), 21-31. Available from: doi:10.1139/g93-004.

ECseq Bioinformatics. (2017) *Why does the per base sequence quality decrease over the read in Illumina?* Available from: https://www.ecseq.com/support/ngs/why-does-the-sequence-quality-decrease-over-the-read-in-illumina [Accessed 22nd November 2019].

El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., Alaux, M., Quesneville, H., Pont, C. & Salse, J. (2017) Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytologist.* 213, 1477-1486. Available from: doi:10.1111/nph.14113.

Eng, J. K., McCormack, A. L. & Yates, J. R., III. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry.* 5 (11), 976-989. Available from: doi:10.1016/1044-0305(94)80016-2.

European Commission. (2019) *EU Plant Variety Database*. Available from: http://ec.europa.eu/food/plant/plant_propagation_material/plant_variety_catalogues_databases/search/public/index.cfm [Accessed 12th September 2019].

Farré, A., Sayers, L., Leverington-Waite, M., Goram, R., Orford, S., Wingen, L., Mumford, C. & Griffiths, S. (2016) Application of a library of near isogenic lines to understand context dependent expression of QTL for grain yield and adaptive traits in bread wheat. *BMC Plant Biology.* 16, 161. Available from: doi:10.1186/s12870-016-0849-6.

Fellers, T. J. & Davidson, M. W. (2019) *Introduction to Confocal Microscopy*. Available from: https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/confocal/confocalintro/ [Accessed 4th December 2019].

Food and Agriculture Organisation of the United Nations. (2017) *FAOSTAT statistics database: Food balance sheets*. Available from: http://www.fao.org/faostat/en/#data/FBS [Accessed 24th September 2019].

Freeman, J., Sparks, C. A., West, J., Shewry, P. R. & Jones, H. D. (2011) Temporal and spatial control of transgene expression using a heat-inducible promoter in transgenic wheat. *Plant Biotechnology Journal.* 9 (7), 788-796. Available from: doi:10.1111/j.1467-7652.2011.00588.x.

Freistroffer, D. V., Kwiatkowski, M., Buckingham, R. H. & Ehrenberg, M. (2000) The accuracy of codon recognition by polypeptide release factors. *Proceedings of the National Academy of Sciences of the United States of America.* 97 (5), 2046-2051. Available from: doi:10.1073/pnas.030541097.

Frietze, S. & Farnham, P. J. (2011) Transcription Factor Effector Domains. In: Hughes, T. (ed.) *A Handbook of Transcription Factors.* Subcellular Biochemistry, vol 52. Dordrecht, Springer, pp. 261-277. Available from: doi:10.1007/978-90-481-9069-0_12.

Ge, H. & Roeder, R. G. (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell.* 78 (3), 513-523. Available from: doi:10.1016/0092-8674(94)90428-6.

Gene Codes Corporation (2017) *Sequencher ® DNA sequence analysis software version 5.4.6.* Ann Arbor, MI, USA, Gene Codes Corporation.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S. & Qi, L. S. (2013a) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 154 (2), 442-451. Available from: doi: 10.1016/j.cell.2013.06.044.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z. R., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S. & Qi, L. S. (2013b) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 154 (2), 442-451. Available from: doi:10.1016/j.cell.2013.06.044.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M. & Weissman, J. S. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell.* 159 (3), 647-661. Available from: doi:10.1016/j.cell.2014.09.029.

Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., Sarah, G., Santoni, S., David, J. & Ranwez, V. (2019) Pervasive hybridizations in the history of wheat relatives. *Science Advances.* 5 (5), eaav9188. Available from: doi:10.1126/sciadv.aav9188.

Glick, B. R. (1995) Metabolic load and heterologous gene expression. *Biotechnology Advances.* 13 (2), 247-261. Available from: doi:10.1016/0734-9750(95)00004-a.

Graham, D. B. & Root, D. E. (2015) Resources for the design of CRISPR gene editing experiments. *Genome Biology.* 16, 260. Available from: doi:10.1186/s13059-015-0823-x.

Grønlund, J. T., Eyres, A., Kumar, S., Buchanan-Wollaston, V. & Gifford, M. L. (2012) Cell specific analysis of *Arabidopsis* leaves using fluorescence activated cell sorting. *Journal of Visualized Experiments*. 68, e4214. Available from: doi:10.3791/4214.

Guadagnuolo, R., Clegg, J. & Ellstrand, N. C. (2006) Relative fitness of transgenic vs. non-transgenic maize x teosinte hybrids: A field evaluation. *Ecological Applications.* 16 (5), 1967-1974. Available from: doi:10.1890/1051-0761(2006)016[1967:RFOTVN]2.0.CO;2.

Guo, M., Davis, D. & Birchler, J. A. (1996) Dosage effects on gene expression in a maize ploidy series. *Genetics.* 142 (4), 1349-1355. Available from: https://www.genetics.org/content/142/4/1349.long [Accessed 1st April 2020].

Guo, Z., Song, Y., Zhou, R., Ren, Z., Jia, J. (2010) Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytologist.* 185 (3), 841-851. Available from: doi:10.1111/j.1469-8137.2009.03099.x.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology.* 166 (4), 557-580. Available from: doi:10.1016/S0022-2836(83)80284-8.

Handy, D. E., Castro, R. & Loscalzo, J. 2011. Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. *Circulation.* 123 (19), 2145-2156. Available from: doi:10.1161/CIRCULATIONAHA.110.956839.

Harmansa, S. & Affolter, M. (2018) Protein binders and their applications in developmental biology. *Development,* 145, dev148874. Available from: doi:10.1242/dev.148874.

Hawkesford, M. J., Araus, J., Park, R., Calderini, D., Miralles, D., Shen, T., Zhang, J. & Parry, M. A. J. (2013) Prospects of doubling global wheat yields. *Food and Energy Security.* 2 (1), 34-48. Available from: doi: 10.1002/fes3.15.

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., Gao, C., Godwin, I. D., Hayes, B. J. & Wulff, B. B. H. (2019) Breeding crops to feed 10 billion. *Nature Biotechnology.* 37, 744-754. Available from: doi:10.1038/s41587-019-0152-9.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E. & Gersbach, C. A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology.* 33 (5), 510-517. Available from: doi:10.1038/nbt.3199.

Hirai, H., Tani, T. & Kikyo, N. (2010) Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *International Journal of Developmental Biology.* 54, 1589-1596. Available from: doi:10.1387/ijdb.103194hh.

Hsu, P. D., Lander, E. S. & Zhang, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* 157 (6), 1262-1278. Available from: 10.1016/j.cell.2014.05.010.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R. & Gornicki, P. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America.* 99 (12), 8133-8138. Available from: doi:10.1073/pnas.072223799.

Huynh, S., Marcussen, T., Felber, F. & Parisod, C. (2019) Hybridization preceded radiation in diploid wheats. *Molecular Phylogenetics and Evolution.* 139, 106554. Available from: doi:10.1016/j.ympev.2019.106554.

Illumina. (2015) *Real Time Analysis (RTA) Software version 2.4.11*. San Diego, CA, USA, Illumina.

Illumina. (2018) *NextSeq Control Software version 2.2.0*. San Diego, CA, USA, Illumina.

Illumina. (2019) *BaseSpace Sequence Hub*. Available from: https://basespace.illumina.com/dashboard [Accessed 29th August 2019].

International Service For The Acquisition Of Agri-Biotech Applications. (2019) *GM approval database*. Available from: http://www.isaaa.org/gmapprovaldatabase/ [Accessed 12th September 2019].

The International Wheat Genome Sequencing Consortium. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science,* 345 (6194), 1251788. Available from: doi:10.1126/science.1251788.

The International Wheat Genome Sequencing Consortium. (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 361 (6403), eaar7191. Available from: doi:10.1126/science.aar7191.

The International Wheat Genome Sequencing Consortium. (2019) *IWGSC RefSeq v2.0 now available at URGI*. Available from: https://www.wheatgenome.org/News/Latest-news/IWGSC-RefSeq-v2.0-now-available-at-URGI [Accessed 22nd September 2019].

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology.* 166 (12), 5429 – 5433. Available from: doi:10.1128/jb.169.12.5429-5433.1987.

Ito, M., Yuan, C., Malik, S., Gu, W., Fondell, J. D., Yamamura, S., Fu, Z., Zhang, X., Qin, J. & Roeder, R. G. (1999) Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Molecular Cell.* 3 (3), 361-370. Available from: doi:10.1016/S1097-2765(00)80463-3.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 337 (6096), 816-821. Available from: doi:10.1126/science.1225829.

Joung, J., Engreitz, J. M., Konnermann, S., Abudayyeh, O. O., Verdine, V. K., Aguet, F., Gootenberg, J. S., Sanjana, N. E., Wright, J. B., Fulco, C. P., Tseng, Y., Yoon, C. H., Boehm, J. S., Lander, E. S. & Zhang, F. (2017) Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature.* 548, 343-346. Available from: doi:10.1038/nature23451.

Kagohara, L. T., Stein-O'Brien, G. L., Kelley, D., Flam, E., Wick, H. C., Danilova, L. V., Easwaran, H., Favorov, A. V., Qian, J., Gaykalova, D. A. & Fertig, E. J. (2018). Epigenetic regulation of gene expression in cancer: Techniques, resources and analysis. *Briefings in Functional Genomics.* 17 (1), 49-63. Available from: doi:10.1093/bfgp/elx018.

Kim, Y., Cha, J. & Chandrasegaran, S. (1996) Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America.* 93 (3), 1156-1160. Available from: doi:10.1073/pnas.93.3.1156.

Kiseleva, A. A., Potokina, E. K. & Salina, E. A. (2017) Features of *Ppd-B1* expression regulation and their impact on the flowering time of wheat near-isogenic lines. *BMC Plant Biology.* 17, 172. Available from: doi:10.1186/s12870-017-1126-z.

Kobayashi, N., Boyer, T. G. & Berk, A. J. (1995) A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Molecular and Cellular Biology.* 15 (11), 6465-6473. Available from: doi:10.1128/MCB.15.11.6465.

Kobayashi-Uehara, A., Shimosaka, E. & Handa, H. (2001) Cloning and expression analyses of cDNA encoding an ADP-ribosylation factor from wheat: tissue-specific expression of wheat ARF. *Plant Science.* 160 (3), 535-542. Available from: doi:10.1016/S0168-9452(00)00416-7.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O. & Zhang, F. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* 517, 583-588. Available from: doi:10.1038/nature14136.

Kotwica-Rolinska, J., Chodakova, L., Chvalova, D., Kristofova, L., Fenclova, I., Provaznik, J., Bertolutti, M., Wu, B. C. & Dolezel, D. (2019) CRISPR/Cas9 genome editing introduction and optimization in the non-model insect *Pyrrhocoris apterus*. *Frontiers in Physiology*. 10, 891. Available from: doi: 10.3389/fphys.2019.00891.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S. & Qi, L. S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols.* 8 (11), 2180-2196. Available from: doi:10.1038/nprot.2013.132.

Leica Microsystems (n.d.) *Leica Application Suite version 4.4*. Wetzlar, Germany, Leica Microsystems.

Li, L., Liu, B., Olsen, K. M. & Wendel, J. F. (2015a) A re-evaluation of the homoploid hybrid origin of *Aegilops tauschii*, the donor of the wheat D-subgenome. *New Phytologist.* 208 (1), 4-8. Available from: doi:10.1111/nph.13294.

Li, L., Liu, B., Olsen, K. M. & Wendel, J. F. (2015b) Multiple rounds of ancient and recent hybridizations have occurred within the *Aegilops-Triticum* complex. *New Phytologist.* 208 (1), 11-12. Available from: doi:10.1111/nph.13563.

Li, Z., Zhang, D., Xiong, X., Yan, B., Xie, W., Sheen, J. & Li, J. (2017) A potent Cas9-derived gene activator for plant and mammalian cells. *Nature Plants.* 3, 930-936. Available from: doi:10.1038/s41477-017-0046-0.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y. & Gao, C. (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications.* 8, 14261. Available from: doi:10.1038/ncomms14261.

Liao, Y., Smyth, G. K. & Shi, W. (2014) featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 30 (7), 923-930. Available from: doi:10.1093/bioinformatics/btt656.

Lin, Y., Ha, I., Maldonado, E., Reinberg, D. & Green, M. R. (1991) Binding of general transcription factor TFIIB to an acidic activating region. *Nature.* 353, 569-571. Available from: doi:10.1038/353569a0.

Lobell, D. B., Schlenker, W. & Costa-Roberts, J. (2011) Climate trends and global crop production since 1980. *Science.* 333 (6042), 616-620. Available from: doi:10.1126/science.1204531.

Love, M. I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.* 15, 550. Available from: doi:10.1186/s13059-014-0550-8.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H. & Joung, J. K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nature Methods.* 10 (10), 977-979. Available from: doi:10.1038/nmeth.2598.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J. & Koonin, E. V. (2011) Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology.* 9, 467-477. Available from: doi:10.1038/nrmicro2577.

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., The International Wheat Genome Sequencing Consortium, Jakobsen, K. S., Wulff, B. B. H., Steuernagel, B., Mayer, K. F. X. & Olsen, O. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science.* 345 (6194), 1250092. Available from: doi:10.1126/science.1250092.

Mendel, G. (1865) Experiments in Plant Hybridization. In: Natural History Society of Brünn. *Meeting of the Brünn Natural History Society*. Brünn, Czech Republic. 1865. pp. 3-47.

Miki, Y., Yoshida, K., Mizuno, N., Nasuda, S., Sato, K. & Takumi, S. (2019) Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*. *DNA Research.* 26 (2), 171-182. Available from: doi:10.1093/dnares/dsy047.

Mittler, G., Stühler, T., Santolin, L., Uhlmann, T., Kremmer, E., Lottspeich, F., Berti, L. & Meisterernst, M. (2003) A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *The EMBO Journal.* 22, 6494-6504. Available from: doi:10.1093/emboj/cdg619.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashitna, H., Hata, K., Nakashima, K. & Hatada, I. (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nature Biotechnology.* 34 (10), 1060-1065. Available from: doi:10.1038/nbt.3658.

Papikian, A., Liu, W., Gallego-Bartolomé, J. & Jacobsen, S. E. (2019) Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. *Nature Communications.* 10, 729. Available from: doi:10.1038/s41467-019-08736-7.

Peabody, D. S. & Berg, P. (1986) Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Molecular and Cellular Biology.* 6 (7), 2695-2703. Available from: doi:10.1128/mcb.6.7.2695.

Pédelacq, J., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology.* 24 (1), 79-88. Available from: doi:10.1038/nbt1172.

Peremarti, A., Twyman, R. M., Gómez-Galera, S., Naqvi, S., Farré, G., Sabalza, M., Miralpeix, B., Dashevskaya, S., Yuan, D., Ramessar, K., Christou, P., Zhu, C., Bassie, L. & Capell, T. (2010) Promoter diversity in multigene transformation. *Plant Molecular Biology.* 73, 363-378. Available from: doi:10.1007/s11103-010-9628-1.

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E. & Gersbach, C. A. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods.* 10 (10), 973-976. Available from: doi:10.1038/nmeth.2600.

Petersen, G., Seberg, O., Yde, M. & Berthelsen, K. (2006) Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution.* 39 (1), 70-82. Available from: doi:10.1016/j.ympev.2006.01.023.

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. & Mahfouz, M. M. (2015) RNA-guided transcriptional regulation *in planta* via synthetic dCas9-based transcription factors. *Plant biotechnology journal.* 13 (4), 578-589. Available from: doi:10.1111/pbi.12284.

Pingali, P. L. (2012) Green revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America.* 109 (31), 12302-12308. Available from: doi:10.1073/pnas.0912953109.

Plotkin, J. B. & Kudla, G. (2011) Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics.* 12, 32-42. Available from: doi:10.1038/nrg2899.

Polstein, L. R. & Gersbach, C. A. (2015) A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology.* 11, 198-200. Available from: doi:10.1038/nchembio.1753.

Purrington, C. B. & Bergelson, J. (1997) Fitness consequences of genetically engineered herbicide and antibiotic resistance in *Arabidopsis thaliana*. *Genetics.* 145 (3), 807-814. Available from: https://www.genetics.org/content/145/3/807.long [Accessed 4th April 2020].

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. & Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 152 (5), 1173-1183. Available from: doi: 10.1016/j.cell.2013.02.022.

Qiu, S., Adema, C. M. & Lane, T. 2005. A computational study of off-target effects of RNA interference. *Nucleic Acids Research.* 33 (6), 1834-1847. Available from: doi:10.1093/nar/gki324.

Raitskin, O., Schudoma, C., West, A. & Patron, N. J. (2019) Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. *PLOS ONE.* 14 (2), e0211598. Available from: doi:10.1371/journal.pone.0211598.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S. J., Cory, A. T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C. J., Chalhoub, B., Mayer, K. F. X., Benhamed, M., Latrasse, D., Bendahmane, A., International Wheat Genome Sequencing Consortium, Wulff, B. B. H., Appels, R., Tiwari, V., Datla, R., Choulet, F., Pozniak, C. J., Provart, N. J., Sharpe, A. G., Paux, E., Spannagl, M., Bräutigam, A. & Uauy, C. (2018) The transcriptional landscape of polyploid wheat. *Science.* 361 (6403), eaar6089. Available from: doi:10.1126/science.aar6089.

Rehman, L., Su, X., Guo, H., Qi, X. & Cheng, H. (2016) Protoplast transformation as a potential platform for exploring gene function in *Verticillium dahliae*. *BMC Biotechnology.* 16, 57. Available from: doi:10.1186/s12896-016-0287-4.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., Bohn, M. & Melchinger, A. E. (2005) Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics.* 110, 859-864. Available from: doi:10.1007/s00122-004-1881-8.

Rock, J. M., Hopkins, F. F., Chavez, A., Diallo, M., Chase, M. R., Gerrick, E. R., Pritchard, J. R., Church, G. M., Rubin, E. J., Sassetti, C. M., Schnappinger, D. & Fortune, S. M. (2017) Programmable

transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nature Microbiology.* 2, 16274. Available from: doi:10.1038/nmicrobiol.2016.274.

Rooke, L., Byrne, D. & Salgueiro, S. (2000) Marker gene expression driven by the maize ubiquitin promoter in transgenic wheat. *Annals of Applied Biology.* 136 (2), 167-172. Available from: doi:10.1111/j.1744-7348.2000.tb00022.x.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F. & Barro, F. (2018) Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnology Journal.* 16 (4), 902-910. Available from: doi:10.1111/pbi.12837.

Sandve, S. R., Marcussen, T., Mayer, K., Jakobsen, K. S., Heier, L., Steuernagel, B., Wulff, B. B. H. & Olsen, O. A. (2015) Chloroplast phylogeny of *Triticum*/*Aegilops* species is not incongruent with an ancient homoploid hybrid origin of the ancestor of the bread wheat D-genome. *New Phytologist.* 208 (1), 9-10. Available from: doi:10.1111/nph.13487.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M. & Barton, G. J. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA.* 22 (6), 839-851. Available from: doi:10.1261/rna.053959.115.

Selma, S., Bernabé-Orts, J. M., Vazquez-Vilar, M., Diego-Martin, B., Ajenjo, M., Garcia-Carpintero, V., Granell, A. & Orzaez, D. (2019) Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator. *Plant Biotechnology Journal.* 17 (9), 1703-1705. Available from: doi:10.1111/pbi.13138.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J. & Gao, C. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology.* 31 (8), 686-688. Available from: doi:10.1038/nbt.2650.

Shan, Q., Wang, Y., Li, J. & Gao, C. (2014) Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protocols.* 9 (1), 2395-2410. Available from: doi:10.1038/nprot.2014.157.

Shewry, P. R. (2009) Wheat. *Journal of Experimental Botany.* 60 (6), 1537-1553. Available from: doi:10.1093/jxb/erp058.

Shiferaw, B., Smale, M., Braun, H., Duveiller, E., Reynolds, M. & Muricho, G. (2013) Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security.* 5, 291-317. Available from: doi:10.1007/s12571-013-0263-y.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X. & Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science.* 351 (6268), 84-88. Available from: doi:10.1126/science.aad5227.

Stöger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P. & Fischer, R. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Molecular Biology.* 42, 583-590. Available from: doi:10.1023/A:1006301519427.

Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell.* 159 (3), 635-646. Available for: doi:10.1016/j.cell.2014.09.039.

Taylor, J. & Woodcock, S. (2015) A perspective on the future of high-throughput RNAi screening: Will CRISPR cut out the competition or can RNAi help guide the way? *Journal of Biomolecular Screening.* 20 (8), 1040-1051. Available from: doi:10.1177/1087057115590069.

Tebaldi, C. & Lobell, D. (2018) Differences, or lack thereof, in wheat and maize yields under three low-warming scenarios. *Environmental Research Letters.* 13 (6), 065001. Available from: doi:10.1088/1748-9326/aaba48.

Technelysium Pty Ltd. (2018) *Chromas version 2.6.5*. South Brisbane, Australia, Technelysium Pty Ltd.

Thermo Fisher Scientific. (2010) *Xcalibur version 2.1*. Waltham, MA, USA, Thermo Fisher Scientific.

Thermo Fisher Scientific. (2012a) *General recommendations for DNA electrophoresis*. Available from: https://www.thermofisher.com/content/dam/LifeTech/global/brands/Documents/1114/general-recommendations-dna-electrophoresis.pdf [Accessed 8th September 2019].

Thermo Fisher Scientific. (2012b) *Proteome Discoverer version 1.4*. Waltham, MA, USA, Thermo Fisher Scientific.

Tourrière, H., Chebli, K. & Tazi, J. (2002) mRNA degradation machines in eukaryotic cells. *Biochimie.* 84 (8), 821-837. Available from: doi:10.1016/S0300-9084(02)01445-1.

The Uniprot Consortium. (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research.* 47 (D1), D506-D515. Available from: doi:10.1093/nar/gky1049.

United Kingdom Government. (2019) *Importing food: Genetically modified foods*. Available from: https://www.gov.uk/food-safety-as-a-food-distributor/genetically-modified-foods [Accessed 22nd November 2019].

United Nations: Department of Economic and Social Affairs: Population Division (2017) *World population prospects: The 2017 revision, key findings and advance tables.* United Nations. Report number: ESA/P/WP/248.

United Nations Framework Convention On Climate Change. (2015) *Paris agreement*. United Nations.

United States Department of Agriculture Animal and Plant Health Inspection Service. (2012) *Secretary Perdue Issues USDA Statement on Plant Breeding Innovation*. Available from:

https://content.govdelivery.com/ accounts/USDAAPHIS/bulletins/1e599ff [Accessed 22nd November 2019].

Unité de Recherche Génomique Info. (2019) *IWGSC RefSeq Annotations*. Available from: https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations [Accessed 15th November 2019].

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M. & Zoldoš, V. (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research.* 44 (12), 5615-5628. Available from: doi:10.1093/nar/gkw159.

Waltz, E. 2016. Gene-edited CRISPR mushroom escapes US regulation. *Nature.* 532, 293. Available from: doi:10.1038/nature.2016.19754.

Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. (2014a) Genetic screens in human cells using the CRISPR-Cas9 system. *Science.* 343 (6166), 80-84. Available from: doi:10.1126/science.1246981.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. & Qiu, J. (2014b) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology.* 32 (9), 947-951. Available from: doi:10.1038/nbt.2969.

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H. & Akhunov, E. (2018) Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *The CRISPR Journal,* 1 (1), 65-74. Available from: doi:10.1089/crispr.2017.0010.

Wheat Initiative. (2016) *Wheat Initiative*. Available from: https://www.wheatinitiative.org/ [Accessed 17th May 2019].

Wulff, B. B. H. & Dhugga, K. S. (2018) Wheat - the cereal abandoned by GM. *Science.* 361 (6401), 451-452. Available from: doi:10.1126/science.aat5119.

Xiao, H., Pearson, A., Coulombe, B., Truant, R., Zhang, S., Regier, J. L., Triezenberg, S. J., Reinberg, D., Flores, O., Ingles, C. J. & Greenblatt, J. (1994) Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Molecular and Cellular Biology.* 14 (10), 7013-7024. Available from: doi:10.1128/MCB.14.10.7013.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G. & Hu, R. (2016) A CRISPR-based approach for targeted DNA demethylation. *Cell Discovery.* 2, 16009. Available from: doi:10.1038/celldisc.2016.9.

Yamamoto, Y. Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K. & Abe, T. (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics.* 8, 67. Available from: doi:10.1186/1471-2164-8-67.

Ying, Y., Yang, X., Zhao, K., Mao, J., Kuang, Y., Wang, Z., Sun, R. & Fei, J. (2015) The Krüppel-associated box repressor domain induces reversible and irreversible regulation of endogenous mouse genes by mediating different chromatin states. *Nucleic Acids Research.* 43 (3), 1549-1561. Available from: doi:10.1093/nar/gkv016.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V. & Zhang, F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell.* 163 (3), 759-771. Available from: doi:10.1016/j.cell.2015.09.038.

Zhang, J., Li, X., Li, G., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y., Xu, J., Chun, N., Yuan, W., Cheng, T. & Zhang, X. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology,* 18, 35. Available from: doi:10.1186/s13059-017-1164-8.

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao, C. & Wang, D. (2018) Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *The Plant Journal.* 94 (5), 857-866. Available from: doi:10.1111/tpj.13903.

Zhang, Z., Hua, L., Gupta, A., Tricoli, D., Edwards, K. J., Yang, B. & Li, W. (2019) Development of an *Agrobacterium*-delivered CRISPR/Cas9 system for wheat genome editing. *Plant Biotechnology Journal.* 17 (8), 1623-1635. Available from: doi:10.1111/pbi.13088.

Zhou, J., Liu, W. J., Peng, S. W., Sun, X. Y. & Frazer, I. (1999) Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *Journal of Virology.* 73 (6), 4972-4982. Available from: https://jvi.asm.org/content/73/6/4972 [Accessed 3rd April 2020].

Zhu, H., Joliot, V. & Prywes, R. (1994) Role of transcription factor TFIIF in serum response factor-activated transcription. *The Journal of Biological Chemistry.* 269 (5), 3489-3497. Available from: https://www.jbc.org/content/269/5/3489.long [Accessed 30th March 2020].

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J., Wang, D. & Gao, C. (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology.* 35 (5), 438-440. Available from: doi:10.1038/nbt.3811.

APPENDIX

A.1 PLASMID MAPS

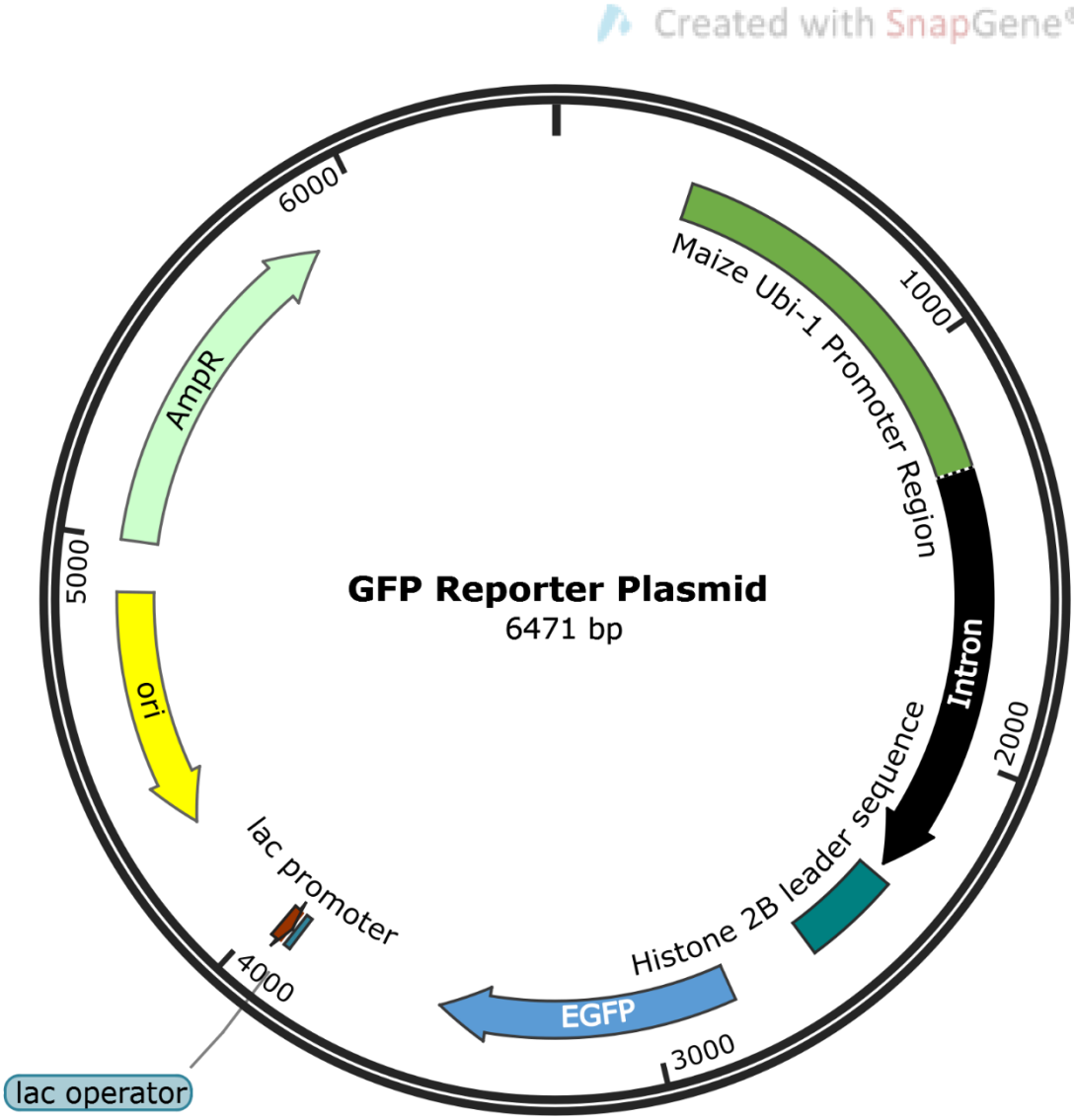

Figure A.1: Map of the GFP reporter plasmid used as a positive control during protoplasting studies. Important features are labelled.

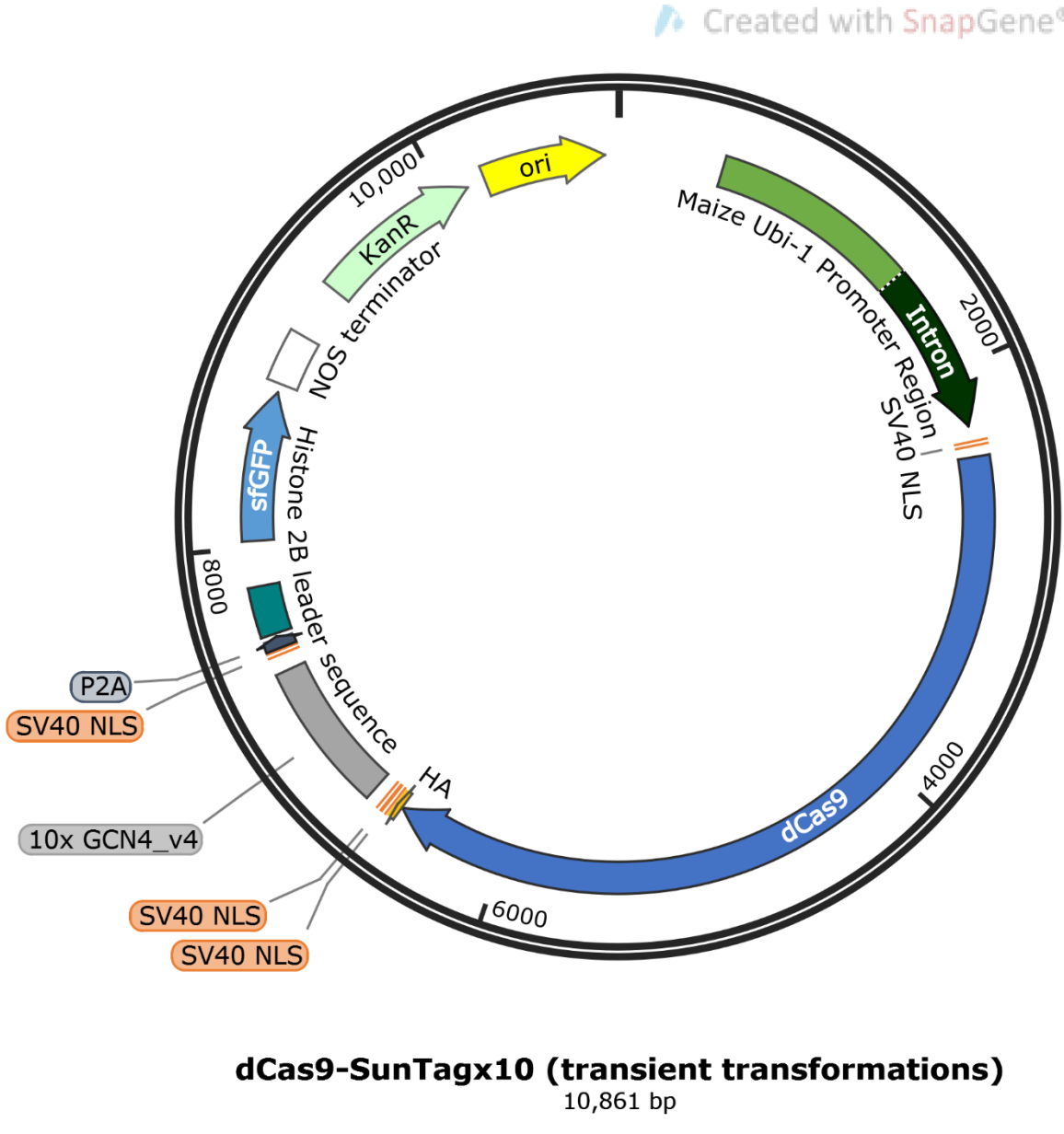

Figure A.2: Map of the dCas9-SunTag$_{x10}$ plasmid used for transient protoplast transformation. Important features are labelled.

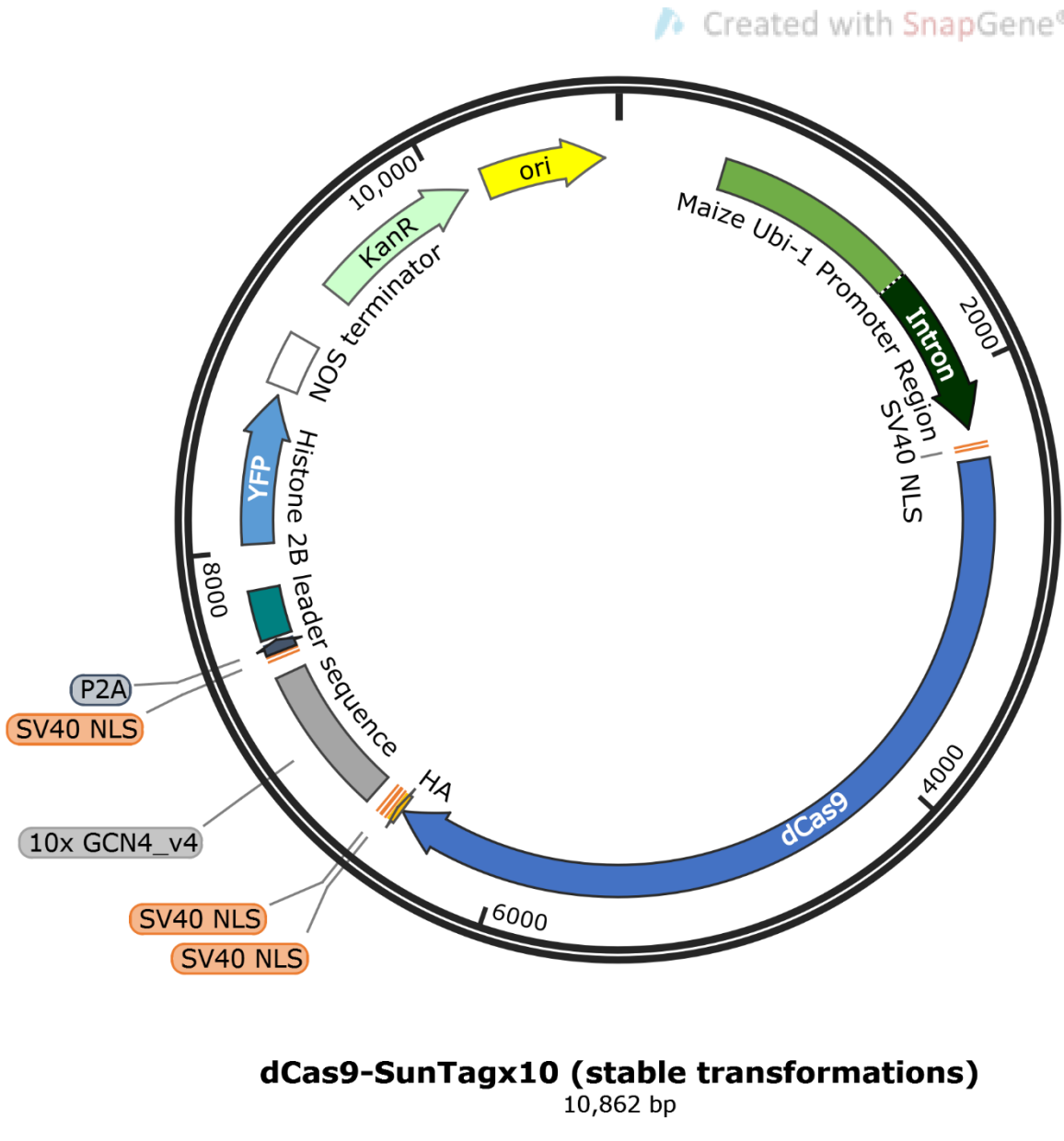

Figure A.3: Map of the dCas9-SunTag$_{x10}$ plasmid used for stable biolistic transformation. Important features are labelled.

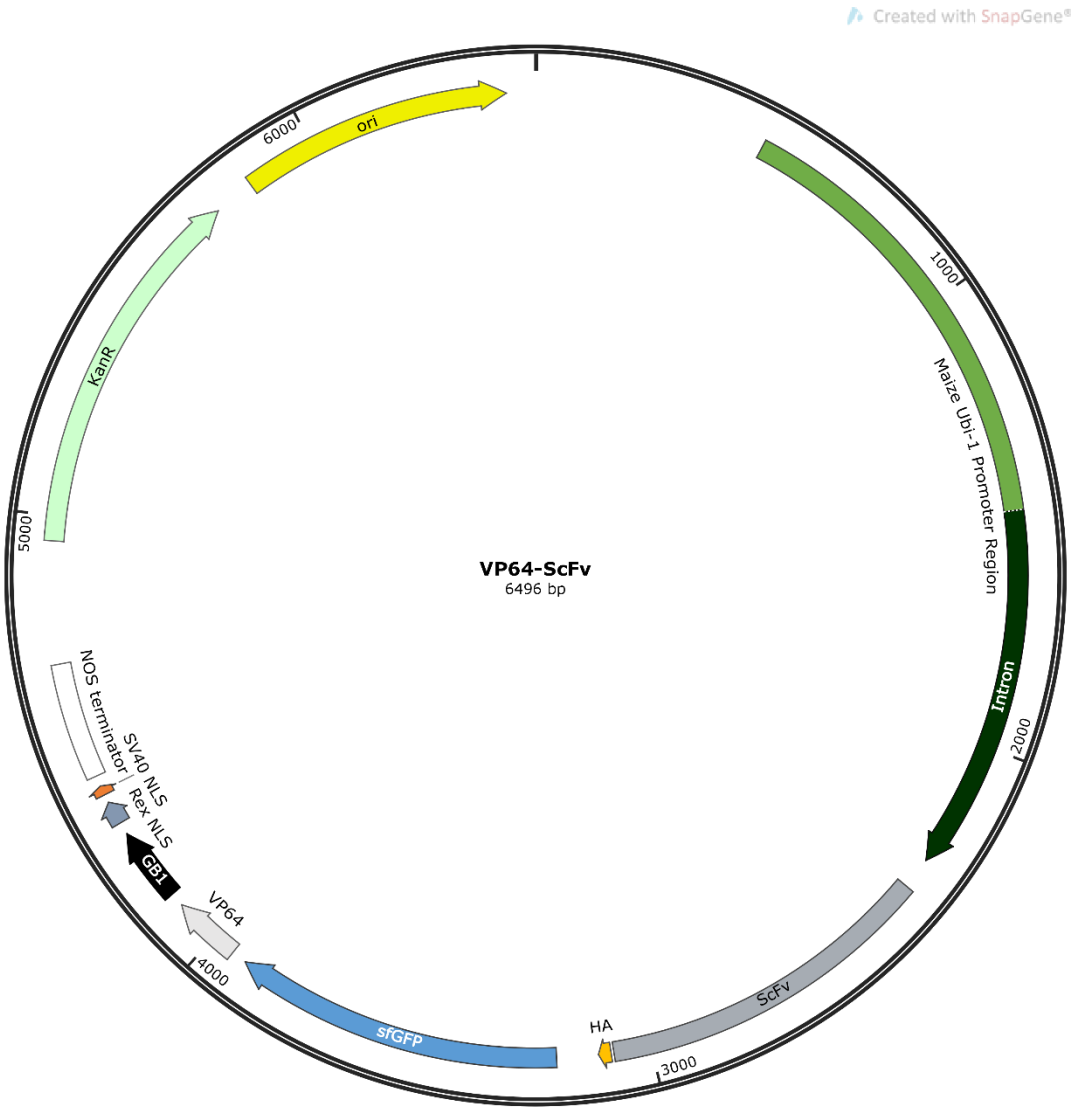

Figure A.4: Map of the VP64-ScFv plasmid used for both transient protoplast transformation and stable biolistic transformation. Important features are labelled.

A.2 CONSTRUCT SEQUENCES

A.2.1 dCas9-SunTag$_{x10}$ (Transient Transformation)

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCATGGATCTCGGGGACGTCTAACTAC
TAAGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGGAAGACTGGGCCTTTCGTTT
TATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGTGAAG
CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAACTAAGCAGAAGGCCAT
CCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTTAGTTAGTTACTTAAGCTCGGGCCCCAAATAAT
GATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTT
GTACAAAAAAGCAGGCTCCACCATGGTACAAAGTGGTGAAGCTTGCATGCCTGCAGTGCAGCGTGACCCGGT
CGTGCCCCTCTCTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTACCACATATTTTTTTTGTCACAC
TTGTTTGAAGTGCAGTTTATCTATCTTTATACATATATTTAAACTTTACTCTACGAATAATATAATCTATAGTACT
ACAATAATATCAGTGTTTTAGAGAATCATATAAATGAACAGTTAGACATGGTCTAAAGGACAATTGAGTATTTT
GACAACAGGACTCTACAGTTTTATCTTTTTAGTGTGCATGTGTTCTCCTTTTTTTTTGCAAATAGCTTCACCTATA
TAATACTTCATCCATTTTATTAGTACATCCATTTAGGGTTTAGGGTTAATGGTTTTTATAGACTAATTTTTTTAGT
ACATCTATTTTATTCTATTTTAGCCTCTAAATTAAGAAAACTAAAACTCTATTTTAGTTTTTTTATTTAATAATTTA
GATATAAAATAGAATAAAATAAAGTGACTAAAAATTAAACAAATACCCTTTAAGAAATTAAAAAAACTAAGGA
AACATTTTTCTTGTTTCGAGTAGATAATGCCAGCCTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCA
GCGAACCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGGCACGGCATCTCTGTCGCTGCCTCTGGACCCC
TCTCGAGAGTTCCGCTCCACCGTTGGACTTGCTCCGCTGTCGGCATCCAGAAATTGCGTGGCGGAGCGGCAGA
CGTGAGCCGGCACGGCAGGCGGCCTCCTCCTCCTCTCACGGCACCGGCAGCTACGGGGGATTCCTTTCCCACC
GCTCCTTCGCTTTCCCTTCCTCGCCCGCCGTAATAAATAGACACCCCCTCCACACCCTCTTTCCCCAACCTCGTGT
TGTTCGGAGCGCACACACACACAACCAGATCTCCCCCAAATCCACCCGTCGGCACCTCCGCTTCAAGGTACGC
CGCTCGTCCTCCCCCCCCCCCCCTCTCTACCTTCTCTAGATCGGCGTTCCGGTCCATGGTTAGGGCCCGGTAGTT
CTACTTCTGTTCATGTTTGTGTTAGATCCGTGTTTGTGTTAGATCCGTGCTGCTAGCGTTCGTACACGGATGCG
ACCTGTACGTCAGACACGTTCTGATTGCTAACTTGCCAGTGTTTCTCTTTGGGGAATCCTGGGATGGCTCTAGC
CGTTCCGCAGACGGGATCGATTTCATGATTTTTTTTGTTTCGTTGCATAGGGTTTGGTTTGCCCTTTTCCTTTATT
TCAATATATGCCGTGCACTTGTTTGTCGGGTCATCTTTTCATGCTTTTTTTTGTCTTGGTTGTGATGATGTGGTG
TGGTTGGGCGGTCGTTCATTCGTTCTAGATCGGAGTAGAATACTGTTTCAAACTACCTGGTGTATTTATTAATT
TTGGAACTGTATGTGTGTGTCATACATCTTCATAGTTACGAGTTTAAGATGGATGGAAATATCGATCTAGGAT
AGGTATACATGTTGATGTGGGTTTTACTGATGCATATACATGATGGCATATGCAGCATCTATTCATATGCTCTA
ACCTTGAGTACCTATCTATTATAATAAACAAGTATGTTTTATAATTATTTTGATCTTGATATACTTGGATGATGG
CATATGCAGCAGCTATATGTGGATTTTTTTAGCCCTGCCTTCATACGCTATTTATTTGCTTGGTACTGTTTCTTTT
GTCGATG:CTC::AC::CCTGTTGTTTGGTGTTACTTCTGCAGGTCGACTCTAGAGGATCTGGAATTCCCGGGGAG
AATCTGGAGCTCTCGAGATTTCTCACGCGTGCCACCATGGGGCCGAAGAAGAAGAGGAAGGTGGGTAGGGT

GTGTCGGATTAGCAGCCTTCGGTACAGGGGTCCTGGCATTGCCACCATGGACAAGAAGTACTCTATCGGGTT
GGCCATAGGAACGAATAGTGTGGGTTGGGCTGTCATCACTGACGAGTACAAGGTGCCATCGAAGAAATTTAA
AGTGCTGGGCAACACCGACAGGCACTCCATCAAGAAGAACCTAATAGGTGCGTTGCTCTTTGATTCCGGGGA
GACAGCCGAAGCGACCCGCCTCAAGCGCACAGCAAGAAGGCGCTATACGCGTAGAAAGAACAGGATCTGCT
ACCTACAAGAGATCTTTTCCAACGAAATGGCGAAAGTGGATGACTCCTTCTTCCATCGGCTCGAGGAAAGCTT
CCTCGTAGAAGAGGACAAGAAGCACGAACGCCATCCGATATTCGGGAACATTGTCGACGAAGTCGCATACCA
CGAGAAGTATCCTACCATCTATCACCTGAGGAAGAAGCTCGTAGATAGCACTGACAAGGCTGATTTGCGCCTC
ATCTATTTGGCCCTGGCCCATATGATTAAGTTCCGTGGGCACTTCCTCATCGAAGGCGACCTTAATCCGGACAA
TTCCGATGTTGACAAACTCTTCATCCAACTGGTCCAGACATACAATCAGCTCTTTGAAGAGAACCCCATCAACG
CGTCAGGAGTTGATGCGAAGGCAATCTTGAGTGCTCGGCTGTCGAAAAGCCGGAGGCTCGAAAATCTGATAG
CCCAACTGCCAGGCGAAAAGAAGAACGGATTGTTTGGCAACCTTATTGCCCTTAGTCTAGGGCTCACCCCTAA
CTTCAAGTCTAACTTTGACCTGGCCGAGGATGCCAAGCTCCAACTGTCCAAAGACACGTACGACGACGATCTC
GACAATCTGCTTGCCCAAATCGGCGACCAGTATGCGGACCTTTTCCTTGCAGCGAAGAACTTGTCAGATGCAA
TCCTCTTGTCCGATATTCTGCGCGTTAACACTGAGATTACCAAGGCTCCTCTTTCGGCATCCATGATCAAACGG
TACGACGAACATCACCAGGATCTCACGCTCCTCAAGGCTCTCGTCAGACAACAGCTGCCGGAGAAGTACAAA
GAGATCTTCTTCGATCAATCAAAGAACGGGTATGCTGGGTACATCGACGGTGGCGCGTCACAGGAGGAGTTC
TACAAGTTTATCAAGCCCATTCTCGAGAAGATGGATGGCACTGAAGAGCTCCTGGTTAAGCTCAATCGCGAGG
ATCTACTCCGCAAGCAGCGTACCTTCGACAATGGAAGCATTCCACACCAAATCCATCTGGGCGAGCTTCACGC
CATACTTCGTAGGCAAGAGGACTTTTACCCATTCCTCAAAGACAACCGGGAGAAAATCGAGAAGATCCTCACA
TTCAGAATCCCCTACTATGTGGGCCCATTGGCAAGGGGCAACTCTAGGTTCGCGTGGATGACTAGAAAATCG
GAGGAAACGATTACCCCATGGAACTTCGAGGAGGTGGTGGACAAGGGCGCTTCTGCCCAAAGCTTCATCGAG
AGGATGACCAACTTCGACAAGAATCTGCCAAACGAGAAGGTCCTCCCCAAACATTCCCTACTGTACGAGTACT
TCACCGTCTACAACGAGCTGACCAAGGTCAAGTACGTTACTGAGGGTATGAGGAAACCGGCCTTTCTGTCAG
GAGAGCAGAAGAAGGCCATAGTGGATCTCCTTTTCAAGACCAATCGCAAGGTCACAGTCAAGCAGCTAAAGG
AGGACTACTTTAAGAAAATCGAGTGCTTTGACTCTGTGGAGATAAGCGGCGTTGAGGATCGCTTCAACGCCTC
GCTGGGAACGTACCACGACTTGCTCAAGATCATTAAGGATAAAGACTTCCTCGATAACGAAGAGAACGAGGA
CATCTTGGAGGACATCGTTCTGACGCTTACTCTCTTCGAGGACCGCGAGATGATCGAAGAGAGGCTCAAGAC
GTATGCGCACCTCTTTGACGACAAAGTGATGAAGCAGCTGAAGCGCAGGCGCTACACAGGTTGGGGCAGGCT
GAGCAGGAAGCTGATCAATGGCATTCGCGATAAGCAGTCGGGCAAGACCATCCTCGACTTCCTCAAGTCCGA
TGGGTTCGCAAACAGGAACTTTATGCAGCTAATCCACGATGACTCCTTGACCTTCAAGGAGGACATTCAGAAG
GCTCAGGTTTCAGGCCAAGGCGATTCGCTCCACGAGCATATCGCGAATTTGGCCGGATCTCCAGCCATCAAGA
AGGGTATACTTCAAACAGTGAAGGTTGTCGATGAGCTGGTCAAAGTGATGGGTAGACATAAGCCCGAGAATA
TCGTGATTGAGATGGCTCGGGAGAACCAAACCACACAGAAGGGGCAGAAGAACTCCAGGGAAAGGATGAA
GAGAATCGAAGAAGGCATCAAGGAGCTCGGATCACAAATTCTCAAGGAGCATCCGGTCGAAAACACCCAACT

GCAAAACGAGAAGCTCTACCTCTACTACCTCCAAAACGGACGCGATATGTACGTGGACCAGGAGCTCGACATC
AATCGCCTATCTGACTACGATGTCGATGCGATAGTACCACAAAGCTTCCTCAAAGACGACTCTATCGACAACA
AGGTGCTGACCAGGTCCGACAAGAACCGCGGGAAGTCAGACAACGTGCCTTCTGAGGAAGTCGTCAAGAAG
ATGAAGAACTACTGGAGGCAACTCCTCAATGCAAAGCTCATTACACAAAGGAAGTTCGACAACCTTACCAAGG
CGGAGAGAGGCGGCCTCAGTGAGCTAGATAAGGCCGGATTTATCAAGCGCCAACTGGTTGAGACACGGCAA
ATCACCAAGCACGTTGCACAGATCCTGGACTCACGCATGAACACAAAATACGATGAGAATGACAAGCTAATCC
GGGAAGTCAAAGTTATTACTCTGAAATCTAAACTGGTGTCCGACTTTCGGAAGGACTTCCAGTTCTATAAGGT
CCGTGAGATCAACAACTACCATCACGCCCATGATGCCTATCTCAATGCTGTGGTTGGCACTGCACTCATCAAGA
AATATCCCAAGCTTGAGTCGGAGTTCGTCTACGGTGACTACAAGGTTTATGACGTAAGGAAGATGATTGCGA
AGAGTGAGCAGGAAATCGGCAAGGCCACCGCCAAGTACTTCTTTTACAGCAACATAATGAACTTCTTCAAG:A
CTGAGAT:C:ACCCTGGCCAATGGAGAGATCAGAAAACGTCCGCTGATAGAGACTAACGGAGAAACTGGTGA
GATAGTGTGGGACAAGGGGAGAGACTTTGCTACTGTCCGGAAGGTGCTAAGCATGCCACAAGTCAACATCGT
CAAGAAAACGGAAGTCCAAACCGGCGGCTTCTCGAAAGAGTCAATCCTTCCGAAGCGCAATAGCGATAAGCT
TATAGCCCGTAAGAAAGACTGGGATCCAAAGAAGTACGGTGGCTTTGACAGTCCTACAGTGGCGTACTCCGT
GCTGGTTGTTGCCAAGGTGGAAAAGGGGAAGTCGAAGAAGCTGAAGAGCGTAAAGGAGTTGCTCGGGATTA
CCATTATGGAGCGTTCCAGCTTTGAGAAGAATCCCATTGATTTCTTGGAGGCGAAAGGATACAAGGAAGTCAA
GAAAGATCTCATCATAAAGTTGCCGAAATACAGCTTGTTCGAGCTGGAGAATGGGAGAAAGAGGATGCTTGC
CAGCGCAGGCGAGTTGCAAAAGGGGAACGAGCTGGCACTCCCATCCAAGTACGTCAACTTCCTGTATCTCGCC
TCACACTATGAGAAACTTAAGGGTTCCCCAGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAGCAACACAAG
CACTACCTCGATGAGATCATCGAACAGATCAGTGAGTTCTCGAAGAGGGTGATTCTTGCTGACGCTAACCTTG
ACAAGGTGTTGTCCGCCTACAACAAGCATCGGGACAAACCCATCCGGGAGCAGGCAGAAAACATCATACACT
TGTTCACGTTGACGAACCTTGGCGCTCCTGCGGCGTTCAAATATTTTGACACGACGATTGATAGGAAGCGCTA
TACCAGCACCAAAGAAGTACTCGATGCCACCCTAATTCACCAGTCTATTACAGGCCTGTACGAAACGAGGATC
GACTTGTCGCAACTAGGAGGAGATGCCTATCCGTATGACGTGCCAGATTACGCCTCCCTTGGGTCTGGTTCCC
CTAAGAAGAAGCGGAAAGTGGAGGATCCCAAGAAGAAGCGCAAAGTCGACGGCATTGGATCAGGCAGCAAT
GGCTCTTCGGGCAGCAATGGCCCAACTGACGCAGCTGAGGAAGAGCTGCTCTCCAAGAACTACCATCTGGAG
AACGAGGTTGCGAGACTCAAGAAAGGCAGTGGGTCAGGGGAAGAGCTCCTTAGCAAGAACTACCACCTCGA
GAACGAGGTTGCAAGACTCAAGAAGGGGTCTGGGTCCGGAGAAGAGCTGCTCTCGAAGAATTATCACCTGG
AAAACGAAGTCGCGAGACTCAAGAAAGGGTCCGGCTCCGGCGAGGAACTCCTCTCAAAGAACTACCACCTTG
AGAATGAAGTTGCCCGCCTGAAGAAGGGTTCTGGGTCGGGCGAGGAGCTTCTCAGTAAGAATTACCACTTGG
AGAACGAAGTTGCACGGCTTAAGAAGGGCTCAGGTAGTGGCGAGGAGCTACTCAGCAAGAACTATCACCTG
GAGAACGAGGTCGCTCGCCTGAAGAAAGGCAGTGGTAGCGGAGAGGAGCTCCTTTCCAAGAACTACCATCTA
GAAAACGAAGTGGCCAGGCTGAAGAAAGGCTCAGGCTCCGGAGAGGAGTTGCTGTCCAAGAATTACCATCT
GGAGAATGAAGTGGCGCGGTTGAAGAAAGGTAGCGGGAGCGGAGAGGAACTCCTGTCGAAGAATTACCAC

CTCGAGAACGAGGTGGCGAGGCTAAAGAAGGGGTCGGGGAGCGGTGAGGAGCTTCTCTCTAAGAACTACCA
TTTGGAGAACGAGGTAGCCAGGCTCAAGAAGGGTTCAGGGTCGGGCACTGCCGTGAACATAGGAGGCGGG
ACAGGCCCCATGGACCTTCAGAGGCCACAAGGAGGCGGCGGTCCCAAGAAGAAGCGCAAAGTGGGATCCGG
AGCCACGAACTTTTCCCTGCTCAAGCAAGCTGGTGATGTGGAGGAGAATCCTGGCCCAGCTCCAGGGACGTC
CATGGCGAAGGCAGATAAGAAACCAGCGGAGAAGAAACCGGCAGAGAAAACTCCGGCAGTCGAACCAGCA
GCAGCGGCAGAGAAGAAACCAAAAGCCGGAAAGAAACTCCCCAAGGAACCAGCCGTCGCCGGAGACAAGA
AGAAGAAGAGATCAAAGAAGAACGTTGAGACATACAAGATCTACATCTTCAAGGTGTTGAAGCAAGTTCATC
CAGACATCGGAATCTCAAGCAAAGCGATGGGAATTATGAACAGTTTCATCAACGATATTTTTGAGAAGCTCGC
TGGTGAGTCTTCGAAGCTGGCGAGGTACAACAAGAAGCCGACGATTACTTCCAGGGAGATTCAGACTGCGGT
TAGACTTGTGTTGCCCGGTGAGTTGGCGAAACATGCTGTGTCTGAAGGGACTAAGGCGGTTACGAAGTTTAC
GAGTTCTGCTCAGGGCGCCTCTTTCATGGTGTCCAAAGGAGAGGAGCTCTTTACAGGAGTTGTGCCCATTTTG
GTCGAGTTGGATGGCGACGTAAATGGCCACAAGTTCAGCGTGAGAGGCGAAGGTGAAGGTGATGCCACCAA
TGGCAAGCTTACCCTCAAGTTCATCTGTACCACTGGGAAACTGCCGGTTCCATGGCCAACGTTGGTCACCACTC
TCACATATGGTGTTCAGTGCTTTTCGAGGTATCCGGACCACATGAAACGCCATGACTTCTTCAAGTCAGCCATG
CCTGAGGGGTATGTCCAAGAACGCACCATCAGCTTTAAGGACGACGGAACGTACAAGACACGTGCGGAGGTC
AAGTTCGAGGGAGATACACTGGTGAACAGGATCGAACTCAAAGGGATCGACTTCAAAGAGGACGGCAACAT
ACTGGGGCATAAGCTCGAGTACAACTTCAATTCTCACAACGTCTACATTACTGCCGACAAGCAGAAGAACGGC
ATTAAGGCGAACTTCAAAATCCGGCATAATGTGGAGGATGGTAGTGTCCAACTGGCTGACCACTACCAGCAG
AACACGCCGATAGGCGATGGCCCAGTGCTTCTGCCCGACAATCACTACCTGTCAACCCAATCCGTTCTGTCTAA
GGACCCTAACGAAAAGAGGGATCACATGGTGCTTCTGGAGTTTGTAACGGCTGCAGGCATCACTCATGGGAT
GGATGAGCTCTACAAGTGATAACTGCGGCCGCAGATATCACTAGTGAGCTCGAATTTCCCCGATCGTTCAAAC
ATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAAT
TACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCC
GCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTG
TCATCTATGTTACTAGATCGGGAATTAATGATATCTCCGCGGCCGCACTCGAGATATCTAGACCCAGCTTTCTT
GTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAAT
AAAATCATTATTTGCCATCCAGCTGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATA
AAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATT
CAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGC
GATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGA
AACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTA
TGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGGTGATGCATGGTTACTCACCACTGCGATCCCCGGA
AAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCC
TGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGC

AATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC
AAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT
GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGAT
ACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAA
TATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTG
GTTAATTGGTTGTAACATTATTCAGATTGGGCCCCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAG
GATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAA
ATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG
GACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT
GGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTT

A.2.2 dCas9-SunTag$_{x10}$ (Stable Transformation)

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCATGGATCTCGGGGACGTCTAACTAC
TAAGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGGAAGACTGGGCCTTTCGTTT
TATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGTGAAG
CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAACTAAGCAGAAGGCCAT
CCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTTAGTTAGTTACTTAAGCTCGGGCCCCAAATAAT
GATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTT
GTACAAAAAAGCAGGCTCCACCATGGTACAAAGTGGTGAAGCTTGCATGCCTGCAGTGCAGCGTGACCCGGT
CGTGCCCCTCTCTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTACCACATATTTTTTTTGTCACAC
TTGTTTGAAGTGCAGTTTATCTATCTTTATACATATATTTAAACTTTACTCTACGAATAATATAATCTATAGTACT
ACAATAATATCAGTGTTTTAGAGAATCATATAAATGAACAGTTAGACATGGTCTAAAGGACAATTGAGTATTTT
GACAACAGGACTCTACAGTTTTATCTTTTTAGTGTGCATGTGTTCTCCTTTTTTTTTTGCAAATAGCTTCACCTATA
TAATACTTCATCCATTTTATTAGTACATCCATTTAGGGTTTAGGGTTAATGGTTTTTATAGACTAATTTTTTTAGT
ACATCTATTTTATTCTATTTTAGCCTCTAAATTAAGAAAACTAAAACTCTATTTTAGTTTTTTTATTTAATAATTTA
GATATAAAATAGAATAAAATAAAGTGACTAAAAATTAAACAAATACCCTTTAAGAAATTAAAAAAACTAAGGA
AACATTTTTCTTGTTTCGAGTAGATAATGCCAGCCTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCA
GCGAACCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGGCACGGCATCTCTGTCGCTGCCTCTGGACCCC
TCTCGAGAGTTCCGCTCCACCGTTGGACTTGCTCCGCTGTCGGCATCCAGAAATTGCGTGGCGGAGCGGCAGA

CGTGAGCCGGCACGGCAGGCGGCCTCCTCCTCCTCTCACGGCACCGGCAGCTACGGGGGATTCCTTTCCCACC
GCTCCTTCGCTTTCCCTTCCTCGCCCGCCGTAATAAATAGACACCCCCTCCACACCCTCTTTCCCCAACCTCGTGT
TGTTCGGAGCGCACACACACACAACCAGATCTCCCCCAAATCCACCCGTCGGCACCTCCGCTTCAAGGTACGC
CGCTCGTCCTCCCCCCCCCCCCCTCTCTACCTTCTCTAGATCGGCGTTCCGGTCCATGGTTAGGGCCCGGTAGTT
CTACTTCTGTTCATGTTTGTGTTAGATCCGTGTTTGTGTTAGATCCGTGCTGCTAGCGTTCGTACACGGATGCG
ACCTGTACGTCAGACACGTTCTGATTGCTAACTTGCCAGTGTTTCTCTTTGGGGAATCCTGGGATGGCTCTAGC
CGTTCCGCAGACGGGATCGATTTCATGATTTTTTTTGTTTCGTTGCATAGGGTTTGGTTTGCCCTTTTCCTTTATT
TCAATATATGCCGTGCACTTGTTTGTCGGGTCATCTTTTCATGCTTTTTTTTGTCTTGGTTGTGATGATGTGGTG
TGGTTGGGCGGTCGTTCATTCGTTCTAGATCGGAGTAGAATACTGTTTCAAACTACCTGGTGTATTTATTAATT
TTGGAACTGTATGTGTGTGTCATACATCTTCATAGTTACGAGTTTAAGATGGATGGAAATATCGATCTAGGAT
AGGTATACATGTTGATGTGGGTTTTACTGATGCATATACATGATGGCATATGCAGCATCTATTCATATGCTCTA
ACCTTGAGTACCTATCTATTATAATAAACAAGTATGTTTTATAATTATTTTGATCTTGATATACTTGGATGATGG
CATATGCAGCAGCTATATGTGGATTTTTTTAGCCCTGCCTTCATACGCTATTTATTTGCTTGGTACTGTTTCTTTT
GTCGATGCTCACCCTGTTGTTTGGTGTTACTTCTGCAGGTCGACTCTAGAGGATCTGGAATTCCCGGGGAGAA
TCTGGAGCTCTCGAGATTTCTCACGCGTGCCACCATGGGGCCGAAGAAGAAGAGGAAGGTGGGTAGGGTGT
GTCGGATTAGCAGCCTTCGGTACAGGGGTCCTGGCATTGCCACCATGGACAAGAAGTACTCTATCGGGTTGG
CCATAGGAACGAATAGTGTGGGTTGGGCTGTCATCACTGACGAGTACAAGGTGCCATCGAAGAAATTTAAAG
TGCTGGGCAACACCGACAGGCACTCCATCAAGAAGAACCTAATAGGTGCGTTGCTCTTTGATTCCGGGGAGA
CAGCCGAAGCGACCCGCCTCAAGCGCACAGCAAGAAGGCGCTATACGCGTAGAAAGAACAGGATCTGCTACC
TACAAGAGATCTTTTCCAACGAAATGGCGAAAGTGGATGACTCCTTCTTCCATCGGCTCGAGGAAAGCTTCCT
CGTAGAAGAGGACAAGAAGCACGAACGCCATCCGATATTCGGGAACATTGTCGACGAAGTCGCATACCACGA
GAAGTATCCTACCATCTATCACCTGAGGAAGAAGCTCGTAGATAGCACTGACAAGGCTGATTTGCGCCTCATC
TATTTGGCCCTGGCCCATATGATTAAGTTCCGTGGGCACTTCCTCATCGAAGGCGACCTTAATCCGGACAATTC
CGATGTTGACAAACTCTTCATCCAACTGGTCCAGACATACAATCAGCTCTTTGAAGAGAACCCCATCAACGCGT
CAGGAGTTGATGCGAAGGCAATCTTGAGTGCTCGGCTGTCGAAAAGCCGGAGGCTCGAAAATCTGATAGCCC
AACTGCCAGGCGAAAAGAAGAACGGATTGTTTGGCAACCTTATTGCCCTTAGTCTAGGGCTCACCCCTAACTT
CAAGTCTAACTTTGACCTGGCCGAGGATGCCAAGCTCCAACTGTCCAAAGACACGTACGACGACGATCTCGAC
AATCTGCTTGCCCAAATCGGCGACCAGTATGCGGACCTTTTCCTTGCAGCGAAGAACTTGTCAGATGCAATCCT
CTTGTCCGATATTCTGCGCGTTAACACTGAGATTACCAAGGCTCCTCTTTCGGCATCCATGATCAAACGGTACG
ACGAACATCACCAGGATCTCACGCTCCTCAAGGCTCTCGTCAGACAACAGCTGCCGGAGAAGTACAAAGAGA
TCTTCTTCGATCAATCAAAGAACGGGTATGCTGGGTACATCGACGGTGGCGCGTCACAGGAGGAGTTCTACA
AGTTTATCAAGCCCATTCTCGAGAAGATGGATGGCACTGAAGAGCTCCTGGTTAAGCTCAATCGCGAGGATCT
ACTCCGCAAGCAGCGTACCTTCGACAATGGAAGCATTCCACACCAAATCCATCTGGGCGAGCTTCACGCCATA
CTTCGTAGGCAAGAGGACTTTTACCCATTCCTCAAAGACAACCGGGAGAAAATCGAGAAGATCCTCACATTCA

GAATCCCCTACTATGTGGGCCCATTGGCAAGGGGCAACTCTAGGTTCGCGTGGATGACTAGAAAATCGGAGG
AAACGATTACCCCATGG:AACTTCGAGGAGGTGGTGGACAAGGGCGCTTCTGCCCAAAGCTTCATCGAGAGG
ATGACCAACTTCGACAAGAATCTGCCAAACGAGAAGGTCCTCCCCAAACATTCCCTACTGTACGAGTACTTCAC
CGTCTACAACGAGCTGACCAAGGTCAAGTACGTTACTGAGGGTATGAGGAAACCGGCCTTTCTGTCAGGAGA
GCAGAAGAAGGCCATAGTGGATCTCCTTTTCAAGACCAATCGCAAGGTCACAGTCAAGCAGCTAAAGGAGGA
CTACTTTAAGAAAATCGAGTGCTTTGACTCTGTGGAGATAAGCGGCGTTGAGGATCGCTTCAACGCCTCGCTG
GGAACGTACCACGACTTGCTCAAGATCATTAAGGATAAAGACTTCCTCGATAACGAAGAGAACGAGGACATC
TTGGAGGACATCGTTCTGACGCTTACTCTCTTCGAGGACCGCGAGATGATCGAAGAGAGGCTCAAGACGTAT
GCGCACCTCTTTGACGACAAAGTGATGAAGCAGCTGAAGCGCAGGCGCTACACAGGTTGGGGCAGGCTGAG
CAGGAAGCTGATCAATGGCATTCGCGATAAGCAGTCGGGCAAGACCATCCTCGACTTCCTCAAGTCCGATGG
GTTCGCAAACAGGAACTTTATGCAGCTAATCCACGATGACTCCTTGACCTTCAAGGAGGACATTCAGAAGGCT
CAGGTTTCAGGCCAAGGCGATTCGCTCCACGAGCATATCGCGAATTTGGCCGGATCTCCAGCCATCAAGAAG
GGTATACTTCAAACAGTGAAGGTTGTCGATGAGCTGGTCAAAGTGATGGGTAGACATAAGCCCGAGAATATC
GTGATTGAGATGGCTCGGGAGAACCAAACCACACAGAAGGGGCAGAAGAACTCCAGGGAAAGGATGAAGA
GAATCGAAGAAGGCATCAAGGAGCTCGGATCACAAATTCTCAAGGAGCATCCGGTCGAAAACACCCAACTGC
AAAACGAGAAGCTCTACCTCTACTACCTCCAAAACGGACGCGATATGTACGTGGACCAGGAGCTCGACATCAA
TCGCCTATCTGACTACGATGTCGATCATATAGTACCACAAAGCTTCCTCAAAGACGACTCTATCGACAACAAGG
TGCTGACCAGGTCCGACAAGAACCGCGGGAAGTCAGACAACGTGCCTTCTGAGGAAGTCGTCAAGAAGATG
AAGAACTACTGGAGGCAACTCCTCAATGCAAAGCTCATTACACAAAGGAAGTTCGACAACCTTACCAAGGCG
GAGAGAGGCGGCCTCAGTGAGCTAGATAAGGCCGGATTTATCAAGCGCCAACTGGTTGAGACACGGCAAAT
CACCAAGCACGTTGCACAGATCCTGGACTCACGCATGAACACAAAATACGATGAGAATGACAAGCTAATCCG
GGAAGTCAAAGTTATTACTCTGAAATCTAAACTGGTGTCCGACTTTCGGAAGGACTTCCAGTTCTATAAGGTCC
GTGAGATCAACAACTACCATCACGCCCATGATGCCTATCTCAATGCTGTGGTTGGCACTGCACTCATCAAGAA
ATATCCCAAGCTTGAGTCGGAGTTCGTCTACGGTGACTACAAGGTTTATGACGTAAGGAAGATGATTGCGAA
GAGTGAGCAGGAAATCGGCAAGGCCACCGCCAAGTACTTCTTTTACAGCAACATAATGAACTTCTTCAAGACT
GAGATCACCCTGGCCAATGGAGAGATCAGAAAACGTCCGCTGATAGAGACTAACGGAGAAACTGGTGAGAT
AGTGTGGGACAAGGGGAGAGACTTTGCTACTGTCCGGAAGGTGCTAAGCATGCCACAAGTCAACATCGTCAA
GAAAACGGAAGTCCAAACCGGCGGCTTCTCGAAAGAGTCAATCCTTCCGAAGCGCAATAGCGATAAGCTTAT
AGCCCGTAAGAAAGACTGGGATCCAAAGAAGTACGGTGGCTTTGACAGTCCTACAGTGGCGTACTCCGTGCT
GGTTGTTGCCAAGGTGGAAAAGGGGAAGTCGAAGAAGCTGAAGAGCGTAAAGGAGTTGCTCGGGATTACCA
TTATGGAGCGTTCCAGCTTTGAGAAGAATCCCATTGATTTCTTGGAGGCGAAAGGATACAAGGAAGTCAAGA
AAGATCTCATCATAAAGTTGCCGAAATACAGCTTGTTCGAGCTGGAGAATGGGAGAAAGAGGATGCTTGCCA
GCGCAGGCGAGTTGCAAAAGGGGAACGAGCTGGCACTCCCATCCAAGTACGTCAACTTCCTGTATCTCGCCTC
ACACTATGAGAAACTTAAGGGTTCCCCAGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAGCAACACAAGCA

CTACCTCGATGAGATCATCGAACAGATCAGTGAGTTCTCGAAGAGGGTGATTCTTGCTGACGCTAACCTTGAC
AAGGTGTTGTCCGCCTACAACAAGCATCGGGACAAACCCATCCGGGAGCAGGCAGAAAACATCATACACTTG
TTCACGTTGACGAACCTTGGCGCTCCTGCGGCGTTCAAATATTTTGACACGACGATTGATAGGAAGCGCTATA
CCAGCACCAAAGAAGTACTCGATGCCACCCTAATTCACCAGTCTATTACAGGCCTGTACGAAACGAGGATCGA
CTTGTCGCAACTAGGAGGAGATGCCTATCCGTATGACGTGCCAGATTACGCCTCCCTTGGGTCTGGTTCCCCTA
AGAAGAAGCGGAAAGTGGAGGATCCCAAGAAGAAGCGCAAAGTCGACGGCATTGGATCAGGCAGCAATGG
CTCTTCGGGCAGCAATGGCCCAACTGACGCAGCTGAGGAAGAGCTGCTCTCCAAGAACTACCATCTGGAGAA
CGAGGTTGCGAGACTCAAGAAAGGCAGTGGGTCAGGGGAAGAGCTCCTTAGCAAGAACTACCACCTCGAGA
AC:GAGGTTGCAAGACTCAAGAAGGGGTCTGGGTCCGGAGAAGAGCTGCTCTCGAAGAATTATCACCTGGAA
AACGAAGTCGCGAGACTCAAGAAAGGGTCCGGCTCCGGCGAGGAACTCCTCTCAAAGAACTACCACCTTGAG
AATGAAGTTGCCCGCCTGAAGAAGGGTTCTGGGTCGGGCGAGGAGCTTCTCAGTAAGAATTACCACTTGGAG
AACGAAGTTGCACGGCTTAAGAAGGGCTCAGGTAGTGGCGAGGAGCTACTCAGCAAGAACTATCACCTGGA
GAACGAGGTCGCTCGCCTGAAGAAAGGCAGTGGTAGCGGAGAGGAGCTCCTTTCCAAGAACTACCATCTAGA
AAACGAAGTGGCCAGGCTGAAGAAAGGCTCAGGCTCCGGAGAGGAGTTGCTGTCCAAGAATTACCATCTGG
AGAATGAAGTGGCGCGGTTGAAGAAAGGTAGCGGGAGCGGAGAGGAACTCCTGTCGAAGAATTACCACCTC
GAGAACGAGGTGGCGAGGCTAAAGAAGGGGTCGGGGAGCGGTGAGGAGCTTCTCTCTAAGAACTACCATTT
GGAGAACGAGGTAGCCAGGCTCAAGAAGGGTTCAGGGTCGGGCACTGCCGTGAACATAGGAGGCGGGACA
GGCCCCATGGACCTTCAGAGGCCACAAGGAGGCGGCGGTCCCAAGAAGAAGCGCAAAGTGGGATCCGGAGC
CACGAACTTTTCCCTGCTCAAGCAAGCTGGTGATGTGGAGGAGAATCCTGGCCCAGCTCCAGGGACGTCCAT
GGCGAAGGCAGATAAGAAACCAGCGGAGAAGAAACCGGCAGAGAAAACTCCGGCAGTCGAACCAGCAGCA
GCGGCAGAGAAGAAACCAAAAGCCGGAAAGAAACTCCCCAAGGAACCAGCCGTCGCCGGAGACAAGAAGA
AGAAGAGATCAAAGAAGAACGTTGAGACATACAAGATCTACATCTTCAAGGTGTTGAAGCAAGTTCATCCAG
ACATCGGAATCTCAAGCAAAGCGATGGGAATTATGAACAGTT:TCATCAACGATATTTTTGAGAAGCTCGCTG
GTGAGTCTTCGAAGCTGGCGAGGTACAACAAGAAGCCGACGATTACTTCCAGGGAGATTCAGACTGCGGTTA
GACTTGTGTTGCCCGGTGAGTTGGCGAAACATGCTGTGTCTGAAGGGACTAAGGCGGTTACGAAGTTTACGA
GTTCTGCTCAGGGCGCCTCTTTCATGGTGTCAAAAGGCGAAGAGCTGTTCACCGGGGTTGTACCCATCCTCGT
CGAGTTGGATGGCGACGTGAATGGCCACAAGTTCAGTGTCTCTGGTGAGGGTGAAGGAGATGCGACATATG
GCAAACTTACCCTCAAGCTGATCTGCACTACGGGCAAATTGCCGGTACCATGGCCTACCCTCGTCACCACACTA
GGGTATGGTGTGCAGTGTTTTGCCAGATACCCAGACCACATGAAGCAGCATGACTTCTTCAAGTCCGCTATGC
CGGAAGGATACGTGCAAGAGAGGACGATCTTCTTTAAGGACGATGGGAACTACAAGACCCGTGCAGAAGTC
AAGTTTGAGGGCGACACACTCGTGAACCGGATTGAGCTCAAGGGCATCGACTTCAAGGAGGACGGGAACATT
CTTGGCCATAAGCTGGAGTACAACTACAATTCGCACAACGTCTACATAACCGCCGACAAGCAGAAGAACGGG
ATTAAGGCCAACTTCAAGATCCGCCATAACATCGAGGATGGAGGTGTTCAATTGGCGGATCACTATCAGCAGA
ATACGCCTATAGGCGATGGTCCCGTTCTACTCCCGGACAACCACTACCTCTCCTACCAAAGCAAGCTGAGCAA

AGACCCCAATGAGAAGCGCGATCACATGGTTCTGCTGGAGTTCGTGACTGCTGCCGGAATCACGCTTGGCAT
GGACGAGCTCTACAAGTGATAACTGCGGCCGCAGATATCACTAGTGAGCTCGAATTTCCCCGATCGTTCAAAC
ATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAAT
TACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCC
GCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTG
TCATCTATGTTACTAGATCGGGAATTAATGATATCTCCGCGGCCGCACTCGAGATATCTAGACCCAGCTTTCTT
GTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAAT
AAAATCATTATTTGCCATCCAGCTGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATA
AAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATT
CAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGC
GATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGA
AACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTA
TGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGGTGATGCATGGTTACTCACCACTGCGATCCCCGGA
AAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCC
TGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGC
AATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC
AAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT
GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGAT
ACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAA
TATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTG
GTTAATTGGTTGTAACATTATTCAGATTGGGCCCCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAG
GATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAA
ATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTRATCCTGTTCACCAGTGKMTGCTSCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGC
GGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG
GCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTT

A.2.3 VP64-ScFv

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCATGGATCTCGGGGACGTCTAACTAC
TAAGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGGAAGACTGGGCCTTTCGTTT

TATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGTGAAG
CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAACTAAGCAGAAGGCCAT
CCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTTAGTTAGTTACTTAAGCTCGGGCCCCAAATAAT
GATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTT
GTACAAAAAAGCAGGCTCCACCATGGTACAAAGTGGTGAAGCTTGCATGCCTGCAGTGCAGCGTGACCCGGT
CGTGCCCCTCTCTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTACCACATATTTTTTTTGTCACAC
TTGTTTGAAGTGCAGTTTATCTATCTTTATACATATATTTAAACTTTACTCTACGAATAATATAATCTATAGTACT
ACAATAATATCAGTGTTTTAGAGAATCATATAAATGAACAGTTAGACATGGTCTAAAGGACAATTGAGTATTTT
GACAACAGGACTCTACAGTTTTATCTTTTTAGTGTGCATGTGTTCTCCTTTTTTTTTGCAAATAGCTTCACCTATA
TAATACTTCATCCATTTTATTAGTACATCCATTTAGGGTTTAGGGTTAATGGTTTTTATAGACTAATTTTTTTAGT
ACATCTATTTTATTCTATTTTAGCCTCTAAATTAAGAAAACTAAAACTCTATTTTAGTTTTTTTATTTAATAATTTA
GATATAAAATAGAATAAAATAAAGTGACTAAAAATTAAACAAATACCCTTTAAGAAATTAAAAAAACTAAGGA
AACATTTTTCTTGTTTCGAGTAGATAATGCCAGCCTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCA
GCGAACCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGGCACGGCATCTCTGTCGCTGCCTCTGGACCCC
TCTCGAGAGTTCCGCTCCACCGTTGGACTTGCTCCGCTGTCGGCATCCAGAAATTGCGTGGCGGAGCGGCAGA
CGTGAGCCGGCACGGCAGGCGGCCTCCTCCTCCTCTCACGGCACCGGCAGCTACGGGGGATTCCTTTCCCACC
GCTCCTTCGCTTTCCCTTCCTCGCCCGCCGTAATAAATAGACACCCCCTCCACACCCTCTTTCCCCAACCTCGTGT
TGTTCGGAGCGCACACACACACAACCAGATCTCCCCCAAATCCACCCGTCGGCACCTCCGCTTCAAGGTACGC
CGCTCGTCCTCCCCCCCCCCCCCTCTCTACCTTCTCTAGATCGGCGTTCCGGTCCATGGTTAGGGCCCGGTAGTT
CTACTTCTGTTCATGTTTGTGTTAGATCCGTGTTTGTGTTAGATCCGTGCTGCTAGCGTTCGTACACGGATGCG
ACCTGTACGTCAGACACGTTCTGATTGCTAACTTGCCAGTGTTTCTCTTTGGGGAATCCTGGGATGGCTCTAGC
CGTTCCGCAGACGGGATCGATTTCATGATTTTTTTTGTTTCGTTGCATAGGGTTTGGTTTGCCCTTTTCCTTTATT
TCAATATATGCCGTGCACTTGTTTGTCGGGTCATCTTTTCATGCTTTTTTTTGTCTTGGTTGTGATGATGTGGTG
TGGTTGGGCGGTCGTTCATTCGTTCTAGATCGGAGTAGAATACTGTTTCAAACTACCTGGTGTATTTATTAATT
TTGGAACTGTATGTGTGTGTCATACATCTTCATAGTTACGAGTTTAAGATGGATGGAAATATCGATCTAGGAT
AGGTATACATGTTGATGTGGGTTTTACTGATGCATATACATGATGGCATATGCAGCATCTATTCATATGCTCTA
ACCTTGAGTACCTATCTATTATAATAAACAAGTATGTTTTATAATTATTTTGATCTTGATATACTTGGATGATGG
CATATGCAGCAGCTATATGTGGATTTTTTTAGCCCTGCCTTCATACGCTATTTATTTGCTTGGTACTGTTTCTTTT
GTCGATGCTCACCCTGTTGTTTGGTGTTACTTCTGCAGGTCGA:CTCTAGAGGATCTGGAATTCCCGGGGAGAA
TCTGGAGCTCTCGAGATTTCTCACGCGTGCCACCATGGGGCCAGACATCGTGATGACGCAAAGTCCTAGCTCC
CTCTCTGCTAGCGTCGGTGATAGGGTCACAATCACATGTCGGTCCTCTACAGGCGCTGTGACTACGTCGAACT
ATGCCAGCTGGGTGCAGGAGAAACCCGGCAAACTCTTCAAGGGCCTCATTGGTGGGACGAACAATCGTGCGC
CAGGGGTACCGTCAAGGTTCAGTGGCTCTCTGATCGGAGACAAGGCCACTCTGACCATCTCCTCACTGCAACC
GGAGGACTTTGCCACCTACTTTTGCGCGTTGTGGTACTCCAATCACTGGGTGTTTGGTCAGGGCACTAAGGTG

GAGCTGAAAAGGGGCGGCGGCGGCAGCGGTGGTGGCGGTTCAGGCGGCGGCGGCTCAAGCGGAGGAGGG
TCGGAAGTTAAGCTTCTGGAATCCGGAGGAGGGCTTGTTCAACCCGGAGGTTCCCTAAAGCTCAGCTGCGCA
GTCTCGGGATTCTCTCTTACCGACTATGGGGTCAATTGGGTTCGCCAAGCACCAGGCAGAGGGCTAGAGTGG
ATAGGGGTCATTTGGGGCGATGGGATTACCGACTATAACAGTGCCCTCAAGGATCGGTTCATCATCTCCAAGG
ACAACGGGAAGAACACCGTCTACCTCCAGATGTCGAAGGTTCGCAGCGATGACACCGCGTTGTACTACTGCGT
GACCGGACTCTTCGACTACTGGGGCCAAGGCACGCTTGTGACAGTCTCCAGCTATCCATACGACGTCCCAGAC
TATGCTGGTGGCGGCGGTGGCTCCGGTGGTGGCGGATCAGGCGGAGGCGGATCTGGCGGAGGTGGGAGTC
TAGATCCAGGCGGCGGCGGCTCGGGTTCGAAAGGGGAGGAGCTCTTTACCGGGGTAGTGCCGATCCTTGTC
GAGCTCGATGGTGACGTGAATGGGCATAAGTTCTCGGTAAGAGGGGAAGGGGAAGGCGACGCCACCAATG
GAAAGCTGACCCTAAAGTTCATCTGTACCACCGGGAAACTTCCGGTTCCTTGGCCAACATTGGTGACGACTCT
CACTTACGGAGTTCAGTGCTTCAGTCGGTATCCCGATCACATGAAAAGGCACGACTTCTTCAAGAGCGCCATG
CCAGAGGGGTACGTTCAAGAGAGGACCATCTCCTTTAAGGACGATGGCACCTACAAGACTCGTGCGGAAGTG
AAGTTTGAGGGAGATACACTGGTGAACCGCATCGAACTGAAGGGCATCGACTTTAAGGAGGACGGCAACATT
CTCGGGCATAAGCTCGAGTACAACTTCAACAGCCATAACGTCTACATCACGGCAGACAAGCAGAAGAACGGG
ATTAAGGCCAACTTCAAGATACGCCACAATGTCGAGGACGGATCAGTCCAACTCGCGGATCACTACCAGCAG
AACACTCCCATTGGTGATGGCCCCGTTCTGTTGCCGGACAATCACTACCTGTCTACGCAAAGCGTGCTCTCCAA
AGACCCTAACGAGAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATG
GATGAGCTCTACAAAGGTGGCGGCAGGACTGGTGGTGGTGGCGGTGGAGATGCGCTTGACGACTTCGATCT
AGACATGCTCGGTTCCGATGCCTTGGACGACTTTGACCTGGACATGCTAGGATCGGACGCGTTGGATGACTTC
GACCTTGACATGCTTGGCTCCGATGCGCTGGACGACTTTGATCTCGATATGCTCGGCTCAGGCGGCGGGAGTC
GCACCGAGGAGTACAAGCTGATCCTCAACGGAAAGACGCTCAAGGGCGAGACAACCACGGAAGCCGTAGAC
GCTGCTACCGCCGAGAAGGTGTTCAAGCAGTATGCCAACGACAATGGGGTTGACGGGGAATGGACCTACGA
CGATGCAACCAAGACCTTCACCGTCACGGAAGGAGGCGGGTCTGGAGGCGGGACAAGCCCCAAAACTCGGC
GCAGACCCAGGAGGAGCCAACGCAAACGTCCACCGACTCCATGGCCTGGCGGAGGGCCGAAGAAGAAGAG
GAAGGTGTAATAGTGAGCTCGAATTTCCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCC
TGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATG
CATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACA
AAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTAATGAT
ATCTTTGCGGCCGCACTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCT
TATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGCAGCTCTG
GCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTG
CTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTC
CAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTAT
CGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT

ACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTAC
TCCTGGTGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCT
GATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTG
TCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGA
GTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATT
CTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAG
GTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGG
TGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCA
GTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACATTATTCAGATTGGGCCCC
GTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT
GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACC
ACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG
GAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCT
GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCT
TTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT

A.3 ANTIBODIES, PRIMERS AND sgRNAS

Table A.1: Details of the antibodies used throughout this study. All antibodies were from Genscript Biotech (Piscataway, NJ, USA).

Antibody Name	**Peptide Raised Against**	**Antigen Position on dCas9-SunTag$_{x10}$ Protein (amino acid number)**
Cas9 1	CTNFDKNLPNEKVLP	583-596
Cas9 2	CGFSKESILPKRNSD	1191-1204
Cas9 3	KVPSKKFKVLGNTDC	113-126

Table A.2: Details of the sgRNAs used throughout this study, as supplied by Professor Keith Edwards (University of Bristol; Bristol, UK). All sgRNA sequences were used as part of Btg21 constructs ligated into pUC19 cloning vectors.

sgRNA Name	**Target Gene**	**Oligo Sequence**	**Target Sequence Position (0 = estimated transcription start site)**
Ub 1	Maize *ubi1*	GCGAACCAGCAGCGTCGCGT	-258
Ub 2	Maize *ubi1*	GCCAAGCGAAGCAGACGGCA	-235
Ub 3	Maize *ubi1*	GTCGGCATCCAGAAATTGCG	-148
Ub 4	Maize *ubi1*	GCGGCAGACGTGAGCCGGCA	-121
PPD 1	Wheat *Ppd-D1*	GTCCTCCTCCTCCACCTGAC	+68
PPD 2	Wheat *Ppd-D1*	GCTCTGTTCCTGCTCGATTG	+137
PPD 3	Wheat *Ppd-D1*	GCCGCCGGCCGCTCCATTCG	+315
PPD 4	Wheat *Ppd-D1*	GGGACTCCCGGATGCGACCG	+452
PPD 5	Wheat *Ppd-D1*	GACCCCAACATGTTTCCTCT	+552

Table A.3: Details of the primers used throughout this study. All primers were from Eurofins Genomics. (Ebersberg, Germany).

Primer Name	**Sequence**
Ub (F)	CCAACCTCGTGTTGTTCGGAG
ARF (F)	GGCTCTCCAACAACATTGCCAAC
H2B (R)	CAGACACAGCATGTTTCGCCAACTC
dCas9 (R)	ATGGAGTGCCTGTCGGTGTTG
VP64-ScFv (R)	CCGAGACTGCGCAGCTGAG
ARF (R)	GGCTTCTGCCTGTCACATACGC

A.4 MASS SPECTROMETRY PARAMETERS

- Mode: data-dependent acquisition
- Survey scan resolution: 60 000 at m/z 400
    - Mass range: m/z 300-2000
- Charge state filtering: enabled
- Dynamic exclusion: enabled
    - Repeat duration: 30 s
    - Repeat count: 1
    - Exclusion list size: 500
- Fragmentation conditions
    - Normalized collision energy: 40%
    - Activation q: 0.25
    - Activation time: 10 ms
    - Minimum ion selection intensity: 500 counts

The top 20 multiply charged ions in each duty cycle were used in the LTQ linear ion trap for MS/MS.

A.5 PROTEOME DISCOVERER PARAMETERS

- Peptide precursor mass tolerance: 10 ppm
- MS/MS tolerance: 0.8 Da
- Mode: full tryptic digestion
- Maximum number of missed cleavages: 2
- Reverse database search: enabled
- False discovery rate cut off: 1%

A.6 STAR PARAMETERS

- outFilterScoreMinOverLread: 0.1
- outFilterMatchNminOverLread: 0.1
- outSAMtype: BAM Sorted by coordinate
- runThreadN: 100

A.7 FEATURECOUNTS PARAMETERS

- p (isPairedEnd): enabled
- t (GTF.featureType): exon

- s (isStrandSpecific): 2 (reversely stranded)

A.8 RNA-SEQ SUMMARY DATA

Table A.4: The number of raw reads obtained for each sample in the RNA-seq run.

Sample	**Number of raw reads**
1A	83,907,448
1B	67,648,390
2A	66,363,668
2B	66,842,388
3A	67,894,884
3B	74,915,550
4A	69,834,746
4B	69,570,932
5A	64,663,540
5B	64,046,360
6A	62,720,562
6B	64,839,602

A.9 ADDITIONAL FLUORESCENT MICROSCOPY IMAGES

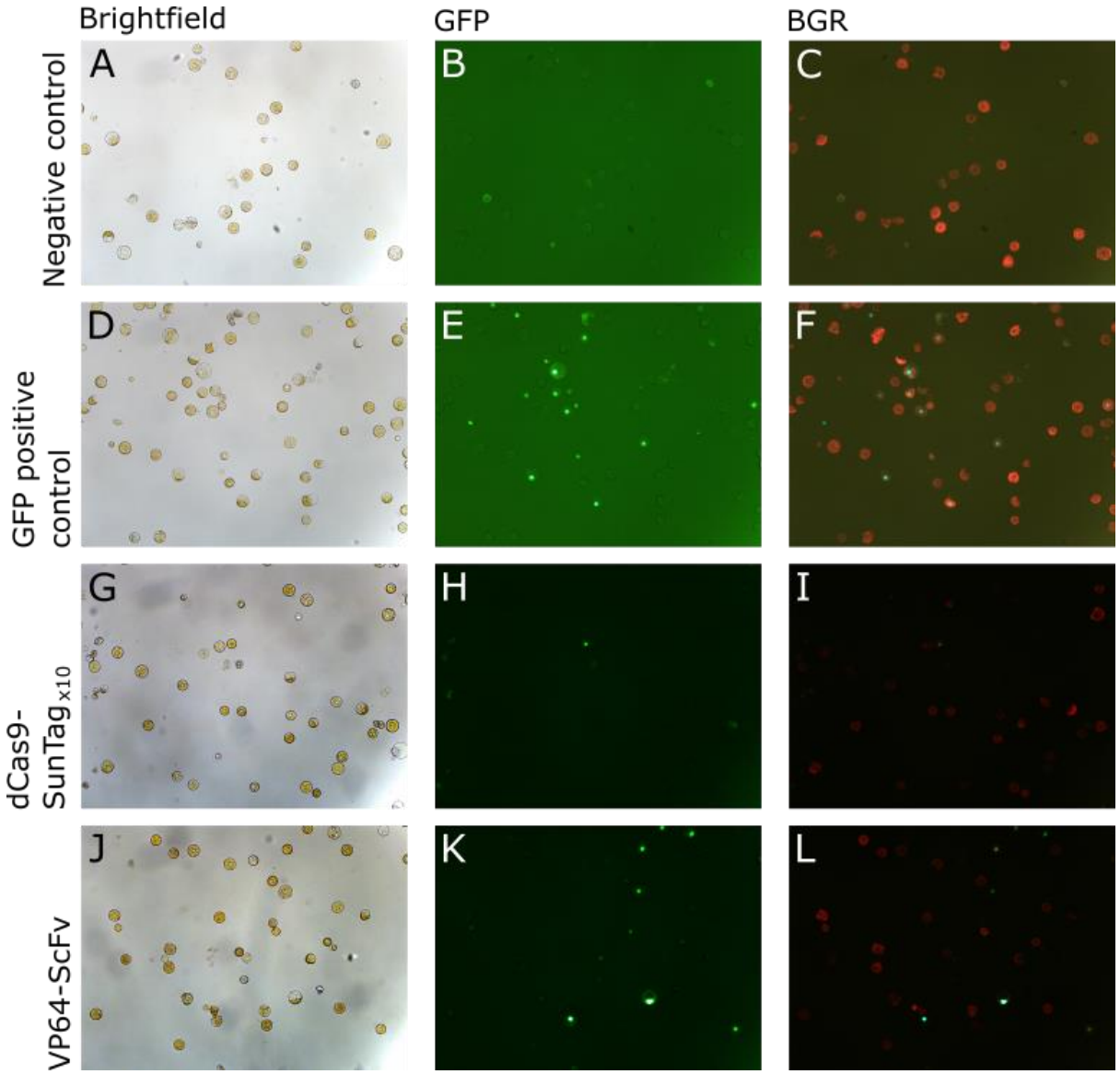

Figure A.5: Fluorescent microscopy images of wheat protoplasts. The protoplasts were incubated for approximately 25 h post transformation. Excitation: 470 nm; excitation intensity: 20%; magnification: 10x. The protoplast treatments are detailed on the left-hand-side of the images. Images A-C are of negative control protoplasts (water replaced any transformation plasmid). Images D-F are of protoplasts transformed with the GFP control plasmid. Images G-I are of protoplasts transformed with the dCas9-SunTag$_{x10}$ plasmid. Images J-L are of protoplasts transformed with the VP64-ScFv plasmid. The filter cubes used are detailed above the images. Images A, D, G and J were taken using a brightfield filter cube. Images B, E, H and K were taken using a GFP filter cube. Images C, F, I and L were taken using a BGR filter cube.